Networks of intergenic long-range enhancers and SNP-RNAs driving castration-resistant phenotype of human prostate cancer

Anna Glinskii , Shuang Ma , Jun Ma , Denise Grant-Lanza , Chang-Uk Lim , Ian Guest , Stewart Sell, Ralph Buttyan, Gennadi Glinsky

Translational and Functional Genomics Laboratory, Department of Pathology and Laboratory Medicine, and Department of Surgery, Division of Urology, Albany Medical College,
Ordway Cancer Center, Ordway Research Institute, Inc., Center for Medical Science
150 New Scotland Ave, Albany, NY 12208
Genlighttechnologies Corporation, Inc., 939 Coast Blvd., Suite 4M, La Jolla, Ca 92037

Email: genlightec@gmail.com

Running title: Intergenic trans-regulatory non-coding SNP-RNAs

Supplement Description
1. Supplemental Text
2. Supplemental Methods
3. Supplemental References
4. Supplemental Figure Legends
5. Supplemental Figures S1-S10
6. Supplemental Tables S1-S10

Supplemental Text

Transcriptome analysis using high-density tiling arrays demonstrated that vast majority of transcripts in human cells is represented by non-coding RNAs, biogenesis and functional significance of which remain obscure (1-7). Whole-genome transcript mapping studies of H. sapiens discovered a genome-scale highly efficient, pervasive transcription of introns and intergenic sequences and documented presence of promoter-associated short RNAs (PASR) and termini-associated short RNAs (TASR; ref. 1-7). Deep sequencing analysis of small RNAs (<200 nucleotides) from human cells revealed that post-transcriptional processing generates a remarkably diverse family of small non-coding RNAs (8). Fine mapping of transcriptional products originated from active promoters of protein-coding genes by RNA polymerase II (RNAPII) identified a population of small non-coding RNAs of 20 to 90 nucleotides in length termed transcription start site-associated RNAs (TSSa-RNAs; ref. 9). This widespread divergent bidirectional transcription at promoters of protein-coding genes may help to maintain a "competence" state of promoter regions poised for subsequent regulation (9). Consistent with the regulatory role of non-coding RNAs, it has been shown that RNAPII transcription of ncRNAs is required for chromatin remodeling at the fission yeast Schizosaccharomyces pombe fbp1(+) locus during transcriptional activation, which indicates that active transcription through the promoter-associated non-coding regions is required to make promoter sequences accessible to RNAPII and transcriptional activators (10).

Recent significant progress in our knowledge of genetic traits which are associated with increased risk of developing multiple human disorders is driven by combined application of intellectual expertise as well as methodological and conceptual principles of genomics and molecular epidemiology to large-scale genome-wide association studies (GWAS) of SNP variations (reviewed in 11-16). A dominant mechanistic approach is based on considerations of the potential effects of SNP variations on protein-coding genes within or near boundaries of which the genetic variants are located. This protein-centric strategy was recently extended to consider the SNP variants residing within boundaries of genes encoding microRNAs and SNPs within microRNA-target sites in mRNAs (reviewed in 11-16). One of notable common outcomes of GWAS is that many disease-linked SNPs identified to date are located within introns or non-genic regions of human genomes which have no direct relations to known protein-coding sequences or microRNA genes. These data suggest that non-canonical mechanisms of phenotype-altering effects of genetic variations may be relevant. A disease phenocode hypothesis has been articulated that is stating that intergenic DNA sequence variations associated with multiple common human disorders may affect phenotypes *in trans* via non-protein-coding SNP sequence-bearing RNA intermediaries (11-14).

Using microarray expression profiling of all known PCGs and microRNA genes, RNA interference (RNAi) and gene transfer approaches, we are dissecting the molecular anatomy and biological consequences of the epigenetic regulatory cross-talk between transRNAs and microRNAs. We identified a set of 4 "stemness" microRNAs (mir-20b; miR-375; miR-205; miR-486) which are up-regulated in the following

experimental and clinical samples: normal human cells engineered to constitutively express *NALP1*-locus autoimmunity transRNAs, human embryonic stem cells, blood-borne human prostate carcinoma (PC) metastasis precursor cells, clinical PC samples, human PC cell lines selected for increased malignant potential in vivo. We documented altered expression patterns of cancer predisposition (CP) transRNAs in cell lines genetically engineered to stably express selected “stemness” microRNAs and **in human cells engineered to constitutively express *NALP1*-locus autoimmunity transRNAs (Supplemental Figure 3).**

References

1. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M. Global identification of human transcribed sequences with genome tiling arrays. Science 2004; 306:2242-6.
2. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007; 447:799-816.
3. Washietl S, Pedersen JS, Korbel JO, Stocsits C, Gruber AR, Hackermüller J, Hertel J, Lindemeyer M, Reiche K, Tanzer A, Ucla C, Wyss C, Antonarakis SE, Denoeud F, Lagarde J, Drenkow J, Kapranov P, Gingeras TR, Guigó R, Snyder M, Gerstein MB, Reymond A, Hofacker IL, Stadler PF. Structured RNAs in the ENCODE selected regions of the human genome. Genome Res 2007; 17: 852-64.
4. Rozowsky JS, Newburger D, Sayward F, Wu J, Jordan G, Korbel JO, Nagalakshmi U, Yang J, Zheng D, Guigó R, Gingeras TR, Weissman S, Miller P, Snyder M, Gerstein MB. The DART classification of unannotated transcription within the ENCODE regions: associating transcription with known and novel loci. Genome Res 2007; 17:732-45.
5. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 2007; 316: 1484-8.
6. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M. What is a gene, post-ENCODE? History and updated definition. Genome Res 2007; 17:669-81.
7. Gingeras TR. Origin of phenotypes: genes and transcripts. Genome Res 2007; 17:682-90.
8. Fejes-Toth K, Sotirova V, Sachidanandam R, Assaf G, Hannon GJ, Kapranov P, Foissac S, Willingham AT, Duttagupta R, Dumais E, Gingeras TR. Affymetrix ENCODE Transcriptome Project; Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature 2009; 457: 1028-32.
9. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. Divergent transcription from active promoters. Science 2008; 322: 1849-51.
10. Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature 2008; 456:130–4.
11. Glinsky G. Phenotype-defining functions of multiple non-coding RNA pathways. Cell Cycle 2008 7:1630-9.

12. Glinsky GV. An SNP-guided microRNA map of fifteen common human disorders identifies a consensus disease phenocode aiming at principal components of the nuclear import pathway. Cell Cycle 2008 7: 2570-83.
13. Glinsky GV. SNP-guided microRNA maps (MirMaps) of 16 common human disorders identify a clinically-accessible therapy reversing transcriptional aberrations of nuclear import and inflammasome pathways. Cell Cycle 2008 7: 3564-76.
14. Glinsky GV. Disease phenocode analysis identifies SNP-guided microRNA maps (MirMaps) associated with human "master" disease genes. Cell Cycle 2008 7: 3680-94.
15. Glinskii, AB, Ma, J, Ma, S, Grant, D, Lim, C, Sell, S, Glinsky, GV. Identification of intergenic trans-regulatory RNAs containing a disease-linked SNP sequence and targeting cell cycle progression/differentiation pathways in multiple common human disorders. Cell Cycle 2009: 3925-42.
16. Glinsky, GV. Genome template switch hypothesis of cancer predisposition. Regulation of MicroRNA Noncoding RNA Expression, Functional Genomics. In: Proceedings of the 100th Annual Meeting of the American Association for Cancer Research; 2009 Apr 18-22; Denver, CO. Philadelphia (PA): AACR; 2009. Abstract nr 2652.
17. Glinsky GV. Human genome connectivity code links disease-associated SNPs, microRNAs and pyknons. Cell Cycle 2009; 8:925-30.

Supplemental Materials and Methods

Disease associated SNP meta-analysis and mapping of genomic coordinates

Primary data sets of SNPs for meta-analysis of genomic coordinates of SNP variations identified in genome-wide association studies (GWAS) of up to 712,253 samples comprising 221,158 disease cases, 322,862 controls, and 168,233 case/control subjects of obesity GWAS were obtained from previously published studies (References in online Supplement). Mapping of the SNP genomic coordinates was performed based on the NCBI release of Human Genome Build 36.3 (reference assembly). Genomic coordinates of the human K4-K36 domains and human lincRNAs are available in the online Supplement data set (7). Genomic coordinates and gene names of the human bivalent domain genes were obtained from the recently published study (8).

Cell Lines

Human BJ1, RWPE1, HME1, U937, and THP-1 cell lines were obtained from ATCC. hTERT-immortalized BJ1 cells were previously described (1).

Tumor Cell Xenografts. Sub-confluent monolayers of control LNCaP or LNCaP-P1 cells were trypsinized and counted, then resuspended in RPMI-1640 medium with 25% charcoal-dextran stripped-fetal bovine serum (CDS-FBS) and 50% (vol/vol) 2 X Matrigel (Millipore, Inc., Billerica, MA). Aliquots (250 µls) containing 10^6 cells were then injected subcutaneously in the flanks of 8-week male NCR Nu/Nu mice (Taconic, Inc.) that had been surgically castrated 10 days prior through a scrotal incision under anesthesia or sham-castrated involving the scrotal incision without removal of the testis. Tumor size was estimated from largest diameter (a), smallest diameter (b) and height (c) measurements with calipers with tumor volume calculated using the formula $V = \pi/6 \times a \times b \times c$ (Dethlefsen, LA, Prewitt, JMS, Mendelsohn, ML Analysis of tumor growth curves, J Natl Cancer Inst, 40: 389-405, 1968). Mice were euthanized at the end of the experiments and tumors were surgically recovered.

Microarray gene expression analysis

Technical and analytical aspects as well as stringent QC and statistical protocols of gene expression analysis experiments was carried-out essentially as described in our published work (2-6). Briefly, the array hybridization and processing, data retrieval and analysis was carried out using standard sets of the Affymetrix equipment, software, and protocols in state-of-the-art Affymetrix microarray core facility. RNA was extracted from cell cultures of two independent biological replicates of each experimental condition and analyzed for sample purity and integrity using BioAnalyzer (Agilent). Expression analysis of 54,675 transcripts was carried out for each sample in duplicates using Affymetrix HG-U133A Plus 2.0 arrays. Data retrieval and analysis was performed using MAS5.0 software and concordant changes of gene expression for each experimental condition were determined at the statistical threshold p value < 0.05 (two-tailed T-test). All microarray analysis data are publicly available coincidently with the date of publication.

microRNA isolation and activity analysis.

microRNA was extracted from adherent cells lysed on culture plates using the *mir*Vana miRNA Isolation kit (Ambion). Homogenized cell lysates were frozen at -80 °C for at least 24 hours prior to microRNA purification. miRNA concentration was checked using

a NanoDrop (Thermo Scientific) before checking quality on a Bioanalyzer (Agilent Technologies).

To assay the activity of microRNAs in transfected cells we used a miRNA Luciferase Reporter Vector (Signosis) specific for microRNA of interest. The target site sequence of the reporter vector is complementary to the miRNA, therefore a decrease in luciferase signal would indicate an increase in microRNA activity. Cells were transfected with the reporter vector using FuGENE 6 Transfection Reagent (Roche); the transfection was allowed to run 48 hours before the cells were lysed using Luciferase Cell Culture Lysis Reagent (Promega). The lysates were read using the FLUOstar OPTIMA system (BMG Lab Technologies), with 20 micro liters of Luciferase Assay Reagent (Promega) injected into each well immediately prior to reading.

miRNA expression analysis.

To analyze a spectrum of miRNA activity in the infected cell lines, we performed qPCR using the TaqMan Human MicroRNA Array v1.0 (Applied Biosystems) run on the 7900HT Fast Real-Time PCR System, fitted with the specific block to run 384-well TaqMan Low Density Arrays (Applied Biosystems). This TaqMan array is distributed on a micro fluidics card, which allows for high reproducibility with minimal error. The array contains 365 different human miRNA assays and two small nucleolar RNAs that function as endogenous controls for data normalization. All miRNA samples were analyzed for quality control and processed at the Functional Genomics Core of the University of Rochester in Rochester, New York. We used the SDS 2.2 software, the platform for the computer interface with the 7900HT PCR System, to generate normalized data, compare samples, and calculate RQ.

Cell Staining and Flow Cytometry

Cells were stained at a concentration of 1X 10^6 cells per 100 ul of HBSS with 2% HICS. Antibodies at appropriate dilutions (CD14-Pacific Blue, Biolegend, Inc; and CD11b-Alexa Fluor® 647, Biolegend, Inc) were added. Staining duration was for 30min with rotation at 4^{o}C. Cells were then washed with staining medium three times and re-suspended in staining medium. The stained specimens were then analyzed using FACSVantage (BD Biosciences, San Diego, CA; http://www.bdbiosciences.com) or FACSAria with either Diva or CellQuest software (BD Biosciences). The cell counter of the flow cytometers was used to determine cell numbers. Cells were collected into HBSS with 2% HICS.

Induced Differentiation of U937 and THP-1 cells

Approximately 2x10^6 U937 or THP-1 cells (5x10^5 cells/ml) in a 25 cm flask were induced to differentiate by treatment with 20 uM PMA (Sigma-Aldrich) for 4 days.

Lentivirus Production and Generation of stably transfected BJ1, U937, RWPE1, LNCap, and THP-1 cells

Allele-specific sense and anti-sense variants of the rs2670660 sequence of 52 nt in length (Figure 1A; nucleotide sequence shown in shaded box) were chemically synthesized and cloned sequentially into pUC57 plasmid by EcoRV (GeneScript Corporation) and pCDH-CMV-MCS-EF1-copGFP plasmid by EcoRI and NotI (SystemsBio). The integrity and molecular identity of the synthetic sequences as well as designed plasmid vectors were monitored by restriction enzyme mapping analysis and direct sequencing. Lentiviruses were generated by co-transfecting pLentiviral vector

with GFP only plasmids (control cultures) or GFP plasmids with synthetic allele-specific 52 nt sequences of the SNP rs2670660 as shown in Figure 1 (see Ref. 4 and Supplemental Material for additional information) and packaging mix (Invitrogen) into 293FT cells using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen), and then target cells were infected with viral supernatant for 24hr. Flow cytometry analysis for GFP expression were performed to confirm the infection and assess the transfection efficiency. Experiments were carried out using cultures with transfection efficiency > 90%.

Luciferase reporter assays

Luciferase reported assays were carried-out to identify allele-specific features of SNP-bearing RNA and DNA sequences. Enhancer/insulator activities of the 2 kb intergenic DNA sequence containing distinct allelic variants of the rs2670660 NLRP1-locus SNP. Luciferase reporter assays to assess the enhancer activity of 2 kb NALP1-locus intergenic region was carried out in RWPE1 and HEK293 cells transiently expressing either control luciferase reporter plasmids or plasmids containing chemically synthesized 2 kb sequences flanking distinct allelic variants of the *NLRP1*-locus intergenic SNP rs2670660, which is positioned exactly in the middle of the enhancer's sequence. Because allelic differences in the luciferase reporter assays tend to be modest, measurements were controlled by multiple replications of all experimental elements of the enhancer assay. We considered the possibility of multiple levels of variations, including plasmid preps, transfection replicates (along with transfection efficiency measurements, such as using the Renilla lucferase co-transfection controls for normalization). At least three independent experiments were carried out for each setting and three replicates of the luciferase assay were performed in each experiment. Results were validated using at least two independent plasmid preps to control for this potentially significant source of variation.

Colony Growth Assay

Cells from sub-confluent cultures (~ 70% confluence) were seeded in triplicates into 6-well plates (100 cells per well), cultured for 2 weeks, and then stained with 0.1% crystal violet for 5 min. Plates were scanned and number of colonies containing > 50 cells was counted.

Protocols for identification of endogenous trans-regulatory small RNAs encoded by the SNP rs2670660

1. Extract small RNA from cells using mirVana isolation kit. ***(protocol A)***
2. Detect if there is DNA contamination by performing PCR using extracted RNA as template and β-actin as primer. ***(protocol B)***
3. Synthesize cDNA from small RNA extracted. ***(protocol C)***
4. Perform first PCR using primer set 2 (GC2F and GC2R). ***(protocol B)***
5. Clean up PCR product and evaluate cleanup PCR product on 1.2% gel. ***(protocol D)***
6. Perform nested PCR using cleanup first PCR product as template and primer set 1 (GC1F and GC1R) and evaluate nested PCR product on 1.2% gel. ***(protocol B)***
7. Cut the DNA band of interest from the gel, extract and purify the DNA for further sequencing analysis. ***(protocol E)***

APPENDIX

Protocol A: Extract small RNA from cells using mirVana isolation kit, AB #1560, 1561

Cell lysis and homogenization

1. Collect culture medium containing cells and spin cells down. Aspirate the culture medium and wash the cells with 4 ml PBS. Spin cells down again.
2. Add 600 µl Lysis/Binding solution and resuspend cell pellet by pipetting up and down.
3. Add 1/10 volume (60 µl) of miRNA Homogenate Additive to the cell lysate, mix well by vortexing and leave it on ice for 10 min. keep at -80ºC overnight.

❖ Starting number of cells: 5×10^6 cells.

Organic extraction

4. Add a volume of Acid-Phenol:Chloroform (equal to the cell lysate before adding Additive).
    - ❖ Be sure to withdraw from the bottom phase in the bottle of Acid-Phenol:Chloroform.
5. Vortex for 30-60 sec to mix and then centrifuge for 5 min at 10,000 ×g at RT to separate the aqueous and organic phases.
6. Carefully remove the aqueous (upper) phase without disturbing the lower phase, and transfer it to a new tube. Note the volume removed.

Final small RNAs isolation

- ❖ Preheat Elute Solution to 95^0C.
- ❖ Put 100% ethanol at RT.

7. Add 1/3 volume of 100% ethanol to the aqueous phase collected. Mix thoroughly by vortexing.
8. Pipet the lysate/ethanol mixture onto a filter cartridge placed on a collection tube. (up to 700 µl can be applied at a time) Centrifuge at 10,000 ×g (usually 10,000 rpm) for 15 sec to pass the mixture through the filter. Transfer the flow-through to a fresh tube and repeat pipetting and centrifuging for excessive mixture.
9. Pool the collected filtrate. Note the total volume of flow-through.
10. Add 2/3 volume RT 100% ethanol to filtrate and mix thoroughly.
11. Pipet the filtrate /ethanol mixture onto a second filter (up to 700 µl can be applied at a time) and centrifuge at 10,000 ×g (usually 10,000 rpm) for 15 sec to pass the mixture through the filter. Discard the flow-through and repeat pipetting and centrifuging for excessive mixture.
12. Apply 700 µl miRNA Wash Solution 1 to the Filter Cartridge and centrifuge for 5-10 sec. Discard the flow-through and replace the Filter Cartridge into the same collection tube.
13. Apply 500 µl Wash Solution 2/3 and centrifuge for 5-10 sec and discard the flow-through. Repeat once with a second 500 µl Wash Solution 2/3.
14. After discarding the flow-through, spin the Filter Cartridge on collection tube for 1 min to remove residual fluid from the filter.

15. Transfer the Filter Cartridge into a fresh Collection tube. Apply 100 µl of pre-heated 95^{0}C Elution Solution to the center of the filter, close the cap and spin at maximum speed for 20-30 sec to elute the RNA. Keep the RNA at -80^{0}C.

Protocol B: PCR

1. In a clean tube on ice, combine PCR reagents to a 25 µl final volume.

	For each 25 µl reaction
Water, RNase-free	___ µl
PCR Buffer (10X)	2.5 µl
PCR Nucleotide Mix (10mM)	0.5 µl
Taq DNA polymerase (50X)	0.5 µl
template	___ µl
Forward primer (10 µM)	1 µl (0.4 µM final conc.)
Reverse primer (10 µM)	1 µl (0.4 µM final conc.)
Total	25 µl

Mix contents of tube by vortexing, collect contents at bottom of tube and store on ice.

2. Thermal cycle

Initial denaturation	95°C	3 min
For cycle 40 or more Denaturation Annealing Extension	 95°C 55°C 72°C	 30 sec 30 sec 1 min (or 1 or 2 min/kb)
Final extension	72°C	3 min
Hold	4°C	

3. Evaluating PCR product on 1.2 % agarose gel.

Protocol C: RT to synthesize cDNA (for 25 µl cDNA product)

1. In the first tube, mix RNA and reverse primer well by votexing at half speed for several sec, spin the tubes in a microcentrifuge at 4^0C and keep on ice. (total 10µl)

	RNA (µl)	Random Primer (50 µM, µl)	Water (µl)
For RNA	___ µl (=1 µg)	1	__µl (= 9– RNA volume)

2. Heat RNA/primer mixture at 75^0C for 5 min and cool on ice for 5 min.
3. Make master mix from the following items:

	each sample (µl)	Total (µl) x
water	6.75	
RT buffer (5X)	5	
PCR nucleotide mix (10mM)	1.25	
RNase inhibitor (4 units/µl)	1	
M-MLV RT (25X)	1	
Total	15	

   Mix contents of tube by votexing, keep on ice 5 min, vortex again, spin the tubes and keep on ice.
4. Add 15 µl mixtures from step 3 into the first tube on ice.
5. Design a thermal cycler program

   Step 1 42^0C, 30 min

   Step 2 95^0C, 5 min

   Step 3 4^0C for storage until removal from the thermal cycler
6. Dilute 25 µl RT products (cDNA) 10 times and use 10 µl diluted cDNA as template in PCR. Aliquot the leftover and keep in -20^0C freezer.

Protocol D: PCR product purification by Montage PCR Centrifugal Filter Devices (Millipore, p36461)

1. Insert the Montage PCR sample reservoir into one of the two vials provided.
2. Pipette 300 µl distilled water or TE buffer into sample reservoir without touching the membrane. Add 100 µl PCR reaction to the reservoir. Seal with attached cap.
3. Place assembly in a compatible centrifuge and counter-balance with a similar device.
4. Spin the Montage PCR unit at 1000 ×g for 15 min. (no longer than 15 min and no greater than 1000 ×g)

5. Remove assembly from centrifuge. Separate vial from sample reservoir. Save filtrate.
6. Place sample reservoir upright into a clean vial and add 20 µl distilled water or TE buffer carefully to the purple end of the reservoir. Avoid touching the membrane surface.
7. Invert the reservoir into a clean vial and spin at 1000 ×g for 2 min.

Protocol E: Extraction and purification of DNA from gel using QIAquick Gel Extraction Kit

- All centrifugation are carried out at 17,900× (13,000rpm) at RT.
- Add 100% ethanol to Buffer PE before use.
- The yellow color of Buffer QG indicates a pH≤7.5.

Procedure

1. Excise the DNA fragment from the agarose gel with a clean, sharp scapel.
2. Weigh the gel slice in a colorless tube. Add 3 volumes of Buffer QG to 1 volume of gel (100mg=100 µl).
3. Incubate at 50^0C for 10 min (or until the gel slice has completely dissolved). Vortex the tube every 2-3 min to mix during the incubation.
4. After the gel slice has dissolved completely, check that the color of the mixture is yellow (similar to Buffer QC without dissolved agarose).
   If the color of the mixture is orange or violet, add 10 µl of 3 M sodium acetate, pH5.0. the color of the mixture will turn to yellow.
5. Add 1 gel volume of isopropanol to the sample and mix.
   Do not centrifuge the sample at this stage.
6. Place a QIAquick spin column in a provided 2 ml collection tube.
7. Apply the sample to the QIAquick column and centrifuge for 1 min.
   The maximum volume of the column reservoir is 800 µl. for the sample volumes of more than 800 µl, simply load and spin again.
8. Discard flow-through and place QIAquick column back in the same collection tube.
9. Add 0.5 ml of Buffer QC to QIAquick column and centrifuge for 1 min to completely remove agarose traces. Discard the flow-through.
10. Add 0.75 ml of Buffer PE to QIAquick column, let it stand for 2-5 min, and centrifuge for 1 min.
11. Discard the flow-through and centrifuge the QIAquick column for an additional 1 min at 17,900 ×g.
12. Place QIAquick column into a clean 1.5 ml microcentrifuge tube.
13. Add 50 µl of water to the center of the QIAquick membrane, stand for 1 min, and centrifuge the column for 1 min.
14. Detect concentration by nanodrop and keep purified DNA at -20^0C.

Statistical and Bioinformatics Analysis

Detailed protocols for data analysis and documentation of the sensitivity, reproducibility, and other aspects of the quantitative statistical microarray analysis using Affymetrix technology have been reported previously (2-6). Forty to sixty percent of the surveyed genes were called present by Affymetrix Microarray Suite version 5.0 software in these experiments. The concordance analysis of differential gene expression across the data sets was performed using Affymetrix MicroDB version 3.0 and DMT version 3.0 software as described previously (2-6). We processed the microarray data using the Affymetrix Microarray Suite version 5.0 software and performed statistical analysis of the expression data set using the Affymetrix MicroDB and Affymetrix DMT software. The Pearson correlation coefficient for individual test samples and the appropriate reference standard were determined using GraphPad Prism version 4.00 software (GraphPad Software). We calculated the significance of the overlap between the lists of differentially-regulated genes by using the hypergeometric distribution test (9).

We analyzed expression profiling data of 697 clinical samples obtained from 185 control subjects and 350 patients diagnosed with 9 common human disorders (Table 16), including Crohn's disease (59 patients), ulcerative colitis (26 patients), rheumatoid arthritis (20 patients), Huntington's disease (17 patients), autism (15 patients), Alzheimer's disease (36 patients), obesity (14 subjects), prostate cancer (64 patients), and breast cancers (99 patients). Microarray data and associated clinical information are available in Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) using the following accession numbers: GDS2601; GDS810; GDS2824; GDS1615; GDS711; GDS1480; GDS2545; GDS1331; GDS1407; GDS3203; GDS2255.

Method References

1. Holt SE, Glinsky VV, Ivanova AB, Glinsky GV. Resistance to apoptosis in human cells conferred by telomerase function and telomere stability. Mol Carcinog. 1999; 25: 241-8.
2. Glinsky, GV. Glinskii, AB. Berezovskaya, O. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. J Clin Invest; 2005; 115: 1503 - 1521.
3. Glinsky GV, Higashiyama T, Glinskii AB. Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. Clin Cancer Res. 2004 10: 2272-2283.
4. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest. 2004 113: 913-923.
5. Glinsky GV, Krones-Herzig A, Glinskii AB, Gebauer G. Microarray analysis of xenograft-derived cancer cell lines representing multiple experimental models of human prostate cancer. Mol Carcinog. 2003 37: 209-221.
6. Glinskii, AB, Ma, J, Ma, S, Grant, D, Lim, C, Sell, S, Glinsky, GV. Identification of intergenic trans-regulatory RNAs containing a disease-linked SNP sequence and targeting cell cycle progression/differentiation pathways in multiple common human disorders. Cell Cycle 2009: 3925-42.
7. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A. 2009 Jul 1. [Epub ahead of print] PMID: 19571010.

8. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 2008; 4: e1000242.
9. Tavazoie, S, Hughes, JD, Campbell, MJ, Cho, RJ, Church, GM. Systematic determination of genetic network architecture. Nat. Genet. 1999. 22:281-285.

SNP Variations Supplemental References

1. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007 447: 661-678.
2. Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 2008 40: 631-7.
3. Haiman CA, Le Marchand L, Yamamato J, Stram DO, Sheng X, Kolonel LN, Wu AH, Reich D, Henderson BE. A common genetic risk factor for colorectal and prostate cancer. Nat Genet 2007 39: 954-6.
4. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008 40: 638-645.
5. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, Gibbons L, Plant D; Wellcome Trust Case Control Consortium, Wilson AG, Marinou I, Morgan A, Emery P; YEAR consortium, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Worthington J. Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. *Hum Mol Genet.* 2008 Apr 22. [Epub ahead of print]
6. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PI, Le JM, Lee HS, Batliwalla F, Li W, Masters SL, Booty MG, Carulli JP, Padyukov L, Alfredsson L, Klareskog L, Chen WV, Amos CI, Criswell LA, Seldin MF, Kastner DL,

Gregersen PK. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med.* 2007 357: 977-986.

7. Plenge RM, Cotsapas C, Davies L, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007 39: 1477-1482.
8. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, Donn R, Symmons D, Hider S, Bruce IN; Wellcome Trust Case Control Consortium, Wilson AG, Marinou I, Morgan A, Emery P; YEAR Consortium, Carter A, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Strachan D, Worthington J. Rheumatoid arthritis association at 6q23. *Nat Genet.* 2007 39: 1431-1433.
9. Wellcome Trust Case Control Consortium; Australo-Anglo-American Spondylitis Consortium (TASC), Burton PR, Clayton DG, Cardon LR, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007 39: 1329-1337.
10. International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN), Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers J, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, Ojwang JO, James JA, Merrill JT, Gilkeson GS, Seldin MF, Yin H, Baechler EC, Li QZ, Wakeland EK, Bruner GR, Kaufman KM, Kelly JA. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet 2008 40: 204-210.

11. Nath SK, Han S, Kim-Howard X, Kelly JA, Viswanathan P, Gilkeson GS, Chen W, Zhu C, McEver RP, Kimberly RP, Alarcón-Riquelme ME, Vyse TJ, Li QZ, Wakeland EK, Merrill JT, James JA, Kaufman KM, Guthridge JM, Harley JB. A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. Nat Genet 2008 40: 152-154.
12. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, Gunnarsson I, Svenungsson E, Sturfelt G, Jönsen A, Truedsson L, Pons-Estel BA, Witte T, D'Alfonso S, Barizzone N, Danieli MG, Gutierrez C, Suarez A, Junker P, Laustrup H, González-Escribano MF, Martin J, Abderrahim H, Alarcón-Riquelme ME. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat Genet 2008 40:211-216.
13. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PV, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapää-Dahlqvist S, Petri M, Manzi S, Seldin MF, Rönnblom L, Syvänen AC, Criswell LA, Gregersen PK, Behrens TW. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med. 2008 358: 900-909.
14. Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, Li G, Adami HO, Hsu FC, Zhu Y, Bälter K, Kader AK, Turner AR, Liu W, Bleecker ER, Meyers DA, Duggan D, Carpten JD, Chang BL, Isaacs WB, Xu J, Grönberg H. Cumulative association of five genetic variants with prostate cancer. N Engl J Med 2008 358: 910-919.
15. Gudmundsson J, Sulem P, Rafnar T, et al. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nat Genet 2008 40: 281-283.

16. Jin Y, Mailloux CM, Gowan K, Riccardi SL, LaBerge G, Bennett DC, Fain PR, Spritz RA. NALP1 in vitiligo-associated multiple autoimmune disease. N Engl J Med 2007 356: 1216-1225.
17. Fisher SA, Tremelling M, Anderson CA, et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet 2008 40: 710-712.
18. Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet 2007; 39:352-8.
19. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007; 447:1087-93.
20. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007; 39:870-4.
21. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007; 39:865-9.
22. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 2008 40: 623-30.
23. Jaeger E, Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet. 2008 40: 26-8.
24. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, Kemp Z, Gorman M,

Chandler I, Papaemmanuil E, Penegar S, Wood W, Sellick G, Qureshi M, Teixeira A, Domingo E, Barclay E, Martin L, Sieber O; CORGI Consortium, Kerr D, Gray R, Peto J, Cazier JB, Tomlinson I, Houlston RS. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 2007 39: 1315-7.

25. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K; CORGI Consortium, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, Cazier JB, Houlston R. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet. 2007 39: 984-8.

26. Gruber SB, Moreno V, Rozek LS, Rennert HS, Lejbkowicz F, Bonner JD, Greenson JK, Giordano TJ, Fearon ER, Rennert G. Genetic Variation in 8q24 Associated with Risk of Colorectal Cancer. Cancer Biol Ther. 2007 6; [Epub ahead of print].

27. Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, Younkin SG, Younkin CS, Younkin LH, Bisceglio GD, Ertekin-Taner N, Crook JE, Dickson DW, Petersen RC, Graff-Radford NR, Younkin SG. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. Nat Genet. 2009 Feb;41(2):192-8.

28. Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, Hogan MF, Schjeide BM, Hooli B, Divito J, Ionita I, Jiang H, Laird N, Moscarillo T, Ohlsen KL, Elliott K, Wang X, Hu-Lince D, Ryder M, Murphy A, Wagner SL, Blacker D, Becker KD, Tanzi RE. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. Am J Hum Genet. 2008 Nov;83(5):623-32.

29. Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Möller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Toulopoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Mühleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeney LA, Franke B; GROUP, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nöthen MM, Peltonen L, Collier DA, St Clair D, Stefansson K. Large recurrent microdeletions associated with schizophrenia. Nature. 2008 Sep 11;455(7210):232-6.

30. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet. 2008 May;40(5):616-22.

31. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Spitz MR, Eisen T, Amos CI, Houlston RS. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet. 2008 Dec;40(12):1407-9.

32. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S, Krokan HE, Skorpen F, Elvestad MB, Vatten L, Njølstad I, Axelsson T, Chen C,

Goodman G, Barnett M, Loomis MM, Lubiñski J, Matyjasik J, Lener M, Oszutowska D, Field J, Liloglou T, Xinarianos G, Cassidy A; EPIC Study, Vineis P, Clavel-Chapelon F, Palli D, Tumino R, Krogh V, Panico S, González CA, Ramón Quirós J, Martínez C, Navarro C, Ardanaz E, Larrañaga N, Kham KT, Key T, Bueno-de-Mesquita HB, Peeters PH, Trichopoulou A, Linseisen J, Boeing H, Hallmans G, Overvad K, Tjønneland A, Kumle M, Riboli E, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P. Lung cancer susceptibility locus at 5p15.33. Nat Genet. 2008 Dec;40(12):1404-6.

33. Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, Balding D, Scott J, Kooner JS. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. Nat Genet. 2008 Jun;40(6):716-8.

34. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, Inouye M, Freathy RM, Attwood AP, Beckmann JS, Berndt SI; Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Jacobs KB, Chanock SJ, Hayes RB, Bergmann S, Bennett AJ, Bingham SA, Bochud M, Brown M, Cauchi S, Connell JM, Cooper C, Smith GD, Day I, Dina C, De S, Dermitzakis ET, Doney AS, Elliott KS, Elliott P, Evans DM, Sadaf Farooqi I, Froguel P, Ghori J, Groves CJ, Gwilliam R, Hadley D, Hall AS, Hattersley AT, Hebebrand J, Heid IM; KORA, Lamina C, Gieger C, Illig T, Meitinger T, Wichmann HE, Herrera B, Hinney A, Hunt SE, Jarvelin MR, Johnson T, Jolley JD, Karpe F, Keniry A, Khaw KT, Luben RN, Mangino M, Marchini J, McArdle WL, McGinnis R, Meyre D, Munroe PB, Morris AD, Ness AR, Neville MJ, Nica AC, Ong KK, O'Rahilly S, Owen KR, Palmer CN, Papadakis K, Potter S, Pouta A, Qi L; Nurses' Health Study, Randall JC, Rayner NW, Ring SM, Sandhu MS, Scherag A, Sims MA,

Song K, Soranzo N, Speliotes EK; Diabetes Genetics Initiative, Syddall HE, Teichmann SA, Timpson NJ, Tobias JH, Uda M; SardiNIA Study, Vogel CI, Wallace C, Waterworth DM, Weedon MN; Wellcome Trust Case Control Consortium, Willer CJ; FUSION, Wraight , Yuan X, Zeggini E, Hirschhorn JN, Strachan DP, Ouwehand WH, Caulfield MJ, Samani NJ, Frayling TM, Vollenweider P, Waeber G, Mooser V, Deloukas P, McCarthy MI, Wareham NJ, Barroso I, Jacobs KB, Chanock SJ, Hayes RB, Lamina C, Gieger C, Illig T, Meitinger T, Wichmann HE, Kraft P, Hankinson SE, Hunter DJ, Hu FB, Lyon HN, Voight BF, Ridderstrale M, Groop L, Scheet P, Sanna S, Abecasis GR, Albai G, Nagaraja R, Schlessinger D, Jackson AU, Tuomilehto J, Collins FS, Boehnke M, Mohlke KL. Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nat Genet. 2008 Jun;40(6):768-75.

35. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, Najjar S, Nagaraja R, Orrú M, Usala G, Dei M, Lai S, Maschio A, Busonero F, Mulas A, Ehret GB, Fink AA, Weder AB, Cooper RS, Galan P, Chakravarti A, Schlessinger D, Cao A, Lakatta E, Abecasis GR. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet. 2007 Jul;3(7):e115.

36. Hinney A, Nguyen TT, Scherag A, Friedel S, Brönner G, Müller TD, Grallert H, Illig T, Wichmann HE, Rief W, Schäfer H, Hebebrand J. Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. PLoS ONE. 2007 Dec 26;2(12):e1361.

37. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer

D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007 May 11;316(5826):889-94.

38. Hunt SC, Stone S, Xin Y, Scherer CA, Magness CL, Iadonato SP, Hopkins PN, Adams TD. Association of the FTO gene with BMI. Obesity (Silver Spring). 2008 Apr;16(4):902-4.

39. Do R, Bailey SD, Desbiens K, Belisle A, Montpetit A, Bouchard C, Pérusse L, Vohl MC, Engert JC. Genetic variants of FTO influence adiposity, insulin sensitivity, leptin levels, and resting metabolic rate in the Quebec Family Study. Diabetes. 2008 Apr;57(4):1147-50.

40. Grant SF, Li M, Bradfield JP, Kim CE, Annaiah K, Santa E, Glessner JT, Casalunovo T, Frackelton EC, Otieno FG, Shaner JL, Smith RM, Imielinski M, Eckert AW, Chiavacci RM, Berkowitz RI, Hakonarson H. Association analysis of the FTO gene with obesity in children of Caucasian and African ancestry reveals a common tagging SNP. PLoS ONE. 2008 Mar 12;3(3):e1746.

41. Al-Attar SA, Pollex RL, Ban MR, Young TK, Bjerregaard P, Anand SS, Yusuf S, Zinman B, Harris SB, Hanley AJ, Connelly PW, Huff MW, Hegele RA. Association between the FTO rs9939609 polymorphism and the metabolic syndrome in a non-Caucasian multi-ethnic sample. Cardiovasc Diabetol. 2008 Mar 13;7:5.

42. Hotta K, Nakata Y, Matsuo T, Kamohara S, Kotani K, Komatsu R, Itoh N, Mineo I, Wada J, Masuzaki H, Yoneda M, Nakajima A, Miyazaki S, Tokunaga K, Kawamoto M, Funahashi T, Hamaguchi K, Yamada K, Hanafusa T, Oikawa S, Yoshimatsu H, Nakao K, Sakata T, Matsuzawa Y, Tanaka K, Kamatani N, Nakamura Y. Variations in the FTO gene are associated with severe obesity in the Japanese. J Hum Genet. 2008;53(6):546-53.

43. Villani AC, Lemire M, Fortin G, Louis E, Silverberg MS, Collette C, Baba N, Libioulle C, Belaiche J, Bitton A, Gaudet D, Cohen A, Langelier D, Fortin PR, Wither JE, Sarfati M, Rutgeerts P, Rioux JD, Vermeire S, Hudson TJ, Franchimont D. Common variants in the NLRP3 region contribute to Crohn's disease susceptibility. Nat Genet. 2009 Jan;41(1):71-6.

Supplemental Figure legends

Supplemental Figure S1. Identification and characterization of intergenic SNP-RNAs containing SNPs associated with increased risk of developing 21 common human disorders.

A, Majority of disease-linked SNPs identified to date in GWAS is located within introns (29%) or non-genic (39%) regions of human genomes which have no direct relations to known protein-coding sequences, microRNAs, or lincRNA genes.

B, Genome-wide chromosome position mapping of identified in this study SNP-RNAs containing SNPs with documented associations with 21 common human disorders.

C, D. Panels C and D show sequence alignments of 14 evolutionary conserved human SNP-RNAs and corresponding sequences of the mouse genome. Positions of disease-linked SNP nucleotides in human SNP-RNA sequences are underlined.

Supplemental Figure S2. Chromatin state map analysis of genomic sequences encoding evolutionary conserved transRNAs reveals a consensus chromatin domain signature comprising histone H3K27Me3, CBP/CREB, EZH2, and POL2 proteins. Chromatin state maps of corresponding human and mouse genome sequences are visualized using the custom tracks of the UCSC Genome Browser. Color-coded horizontal lines depict alignments of DNA sequences derived from Chip-Seq experiments using antibodies against corresponding proteins. Each color-coded horizontal line represents data from independent biological replicates. Note nearly ubiquitous alignments of the evolutionary-conserved transRNA-encoding sequences within binding sites of the histone H3K27Me3, CBP/CREB, EZH2, and POL2 proteins. Positions of disease-linked SNP nucleotides within transRNA-encoding sequences are indicated by arrows and vertical lines. Original experiments describing the corresponding mouse and human genome-wide chromatin state maps were reported elsewhere (5, 6).

Supplemental Figure S3. Chromatin state map analysis of genomic sequences encoding non-conserved human transRNAs reveals a consensus chromatin domain signature comprising histone H3K27Me3 and EZH2 proteins. Chromatin state maps of individual transRNA-encoding genomic sequences in human embryonic stem cells are visualized using the custom tracks of the UCSC Genome Browser. Color-coded horizontal lines depict alignments of DNA sequences derived from Chip-Seq experiments using antibodies against corresponding proteins. Individual zoom-in chromatin state maps aligned to corresponding transRNA-encoding sequences are shown in groups segregated according to phenotypes associated with the corresponding disease-linked SNPs. Chromatin state maps of extended genomic segments are shown when multiple disease-linked SNPs were identified within continuous genomic regions (size of ~100 kb or less) in addition to individual zoom-in chromatin state maps aligned to corresponding transRNA-encoding sequences. Each color-coded horizontal line represents data from independent biological replicates. Note nearly ubiquitous alignments of the positions of disease-linked SNP nucleotides within binding sites of the histone H3K27Me3 and Polycomb proteins (EZH2 and/or RING1B).

Positions of disease-linked SNP nucleotides within transRNA-encoding sequences are indicated by vertical lines. Original experiments describing the corresponding mouse and human genome-wide chromatin state maps were reported elsewhere (6).

Supplemental Figure S4. Chromatin State Signatures of 43 IDAGL defined by ENCODE data and visualized using UCSC Genome Browser (http://genome.ucsc.edu/ENCODE/).

Supplemental Figure S5. Examples of allele-specific secondary structures of identified to date disease-associated transRNAs. Arrows indicate the positions of nucleotides variations which are associated with increased risk of developing corresponding disorders. Note that disease-linked SNPs are often located within the loop structures of transRNAs and transRNAs containing SNP variants have distinct secondary structures. Note that loop sequences of many identified to date SNP sequence-bearing intergenic and intronic transRNAs have 8-11 nt segments (which include SNP nucleotides) identical to primary sequences of microRNAs, which suggest that one of the mechanisms of transRNA bioactivity may be associated with modulation of the biogenesis and functions of selected microRNAs (11-17). Bottom right panel shows alignments of the microRNA target sites in human PC-susceptibility transRNA A21. Individual human microRNAs (short horizontal bars) aligned along the A21 transRNA sequence according to the positions of respective target sites. Single vertical bar marks the position of PC-predisposition SNP. Note that a vast majority of microRNA target sites segregates to the A21 transRNA segment around PC-predisposition SNP and includes SNP nucleotide.

Supplemental Figure S6. Allele-specific effects of *NALP1*-locus transRNAs on expression of distinct classes of non-coding RNAs in human cells.
A, Microarray analysis of human BJ1 cells engineered to stably express distinct allelic variants of the *NALP1*-locus transRNAs reveals allele-specific alterations of expression levels of multiple distinct classes of non-coding RNAs which include snoRNAs and snoRNA-host genes (*SNORD113; SNHG1; SNHG3; SNHG8*), long non-coding RNAs (*MEG3, tncRNA,* and *MALAT1*), microRNAs, microRNA-precursors, and protein-coding microRNA-host genes (*ATAD2; KIAA1199*). Note that changes of expression of intron-residing microRNAs miR-548d (intron of the *ATAD2* gene) and miR-549 (intron of the *KIAA1199* gene) are in good correspondence with allele-specific expression levels of corresponding microRNA-host genes which suggest a coordinated mechanism of regulation.
B, ABI PCR-based screen identifies a statistically significant set of 36 microRNAs expression of which is altered at least 1.5-fold in *NALP1*-locus transRNA-expressing cells compared to control BJ1/EGFP cells and differentially regulated in pathology-linked G-allele-expressing BJ1 cells compared to the ancestral A-allele-expressing cells. Note that 18 of 36 (50%) of these microRNAs are derived from the single microRNA cluster on ~ 200 kb continuous region of 14q32 band of chromosome 14, which suggest that 14q32 cluster microRNAs may be primary molecular targets of the *NALP1*-locus transRNAs.

Supplemental Figure S7. Sequence homology profiling of *NALP1*-locus transRNAs and transRNA-regulated microRNAs and long non-coding RNAs identifies extensive sequence homology/complementarity features.
A, Genomic location (top left figure), secondary structures of 152 nt (bottom left figure) and 52 nt (top right figure) transRNA molecules, and position of the microRNA-target sites along the 152 nt transRNA sequence (bottom right figure) of *NALP1*-locus transRNAs containing SNP rs2670660.
B, Visualization of individual microRNA-target sites within the *NALP1*-locus transRNA molecule. Note that all 36 *NALP1*-locus transRNA-regulated microRNAs (Figure 5, main text) have at least one potential target sites within 152 nt sequence of the *NLAP1*-locus transRNA molecule and many microRNA target sites manifest allele-associated changes of the minimal free energy (mfe) transRNA/microRNA hybridization.
C, D, microRNAs which are differentially regulated in BJ1 cells expressing distinct allelic variants of the *NALP1*-locus transRNAs share multiple sequence identity segments of at least 11 nucleotides in length with sequences of transRNA-regulated *MEG3* (C) and *MALAT1* (D) long non-coding RNAs.

Supplemental Figure S8. Allele affinity model of transRNA-mediated regulation of microRNA expression and activity. A-C, high affinity (low mfe) transRNA alleles facilitate increase abundance levels of corresponding microRNAs. Inverse correlation between allele-specific changes in minimal free energy (mfe) of transRNA/microRNA hybridization and experimentally-defined changes of microRNA expression and activity that is lower mfe values correspond to higher levels of microRNA expression and activity. These relationships are shown for microRNAs the abundance levels of which in human cells are induced (miR-302a; miR-629; miR-548d; miR-200a; miR-627; miR-770-5p) or repressed (miR-133a; miR-20b; miR-205; let-7b) by forced expression of pathology-linked G-allele transRNAs compared to ancestral A-allele-expressing cells. Insert bars show the results Q-PCR analysis of expression of corresponding microRNAs.
D, Luciferase reporter assay of miR-205 and let-7b activities in early-passage RWPE1 cells stably expressing distinct allelic variants of the NALP1-locus transRNAs demonstrates increased activity of both microRNAs in high affinity ancestral A-allele-expressing cells compared to low affinity pathology-linked G-allele-expressing cells.
E, Application of the allele affinity model of transRNA-mediated regulation of microRNA expression and activity to development of the allele equilibrium hypothesis explaining the phenotype-altering effects of transRNAs as the consequence of direct actions on microRNAs abundance and activity and down-stream effects of transRNA-regulated microRNAs on expression of protein-coding genes.

Supplemental Figure S9. Forced expression of transRNA-regulated microRNAs recapitulates transRNA-induced epigenetic and phenotypic features in human cells.
A-E, Epigenetic regulatory “cross-talk” between *NALP1*-locus transRNAs and poto-oncogenic cancer-susceptibility transRNAs and altered expression profiles of cancer susceptibility transRNAs in human cell lines engineered to express selected transRNA-regulated “stemness” microRNAs.

Panels A-D show the results of validation experiments confirming the specificity of expression of selected microRNAs in GFP-tagged cell lines genetically engineered to stably over-express individual microRNAs or desired combinations of 2 or 3 microRNAs of interest using microarray (A) or Q-RT-PCR (B-D) methods. These cell lines are utilized in experiments designed to determine whether selected microRNAs are essential integral components of transRNA regulatory networks.

E, Distinct expression profiles of transRNA-regulated microRNAs and 6 breast cancer (BC)-susceptibility transRNAs (top-right) and 11 prostate cancer (PC)-susceptibility transRNAs (bottom-left) in human cells. Top left panels of bars show the results of the Q-RT-PCR experiments validating increased expression of miR-375, miR-20b, and miR-205 microRNAs in human cells stably expressing ancestral A-allele *NALP1*-locus transRNAs compared to the pathology-linked G-allele-expressing cells. Forced expression of *NALP1*-locus transRNAs erases epigenetic transcriptional memory of mesenchymal BJ1 cells and induces expression of multiple proto-oncogenic transRNAs (red circles). Bottom gels show expression patterns of cancer-susceptibility transRNAs in BJ1 cells genetically engineered to over-express selected transRNA-regulated microRNAs such as the individual “stemness” microRNAs (miR-205; miR-375; miR-20b) or desired combinations of “stemness” microRNAs. Note that increased expression of 8 cancer susceptibility transRNAs marked in boxes (A7; A8; A11; A18; A21; B5; B6; B7) induced by selected transRNA-regulated “stemness” microRNAs recapitulates the pattern of increased expression induced by *NALP1*-locus autoimmunity transRNAs.

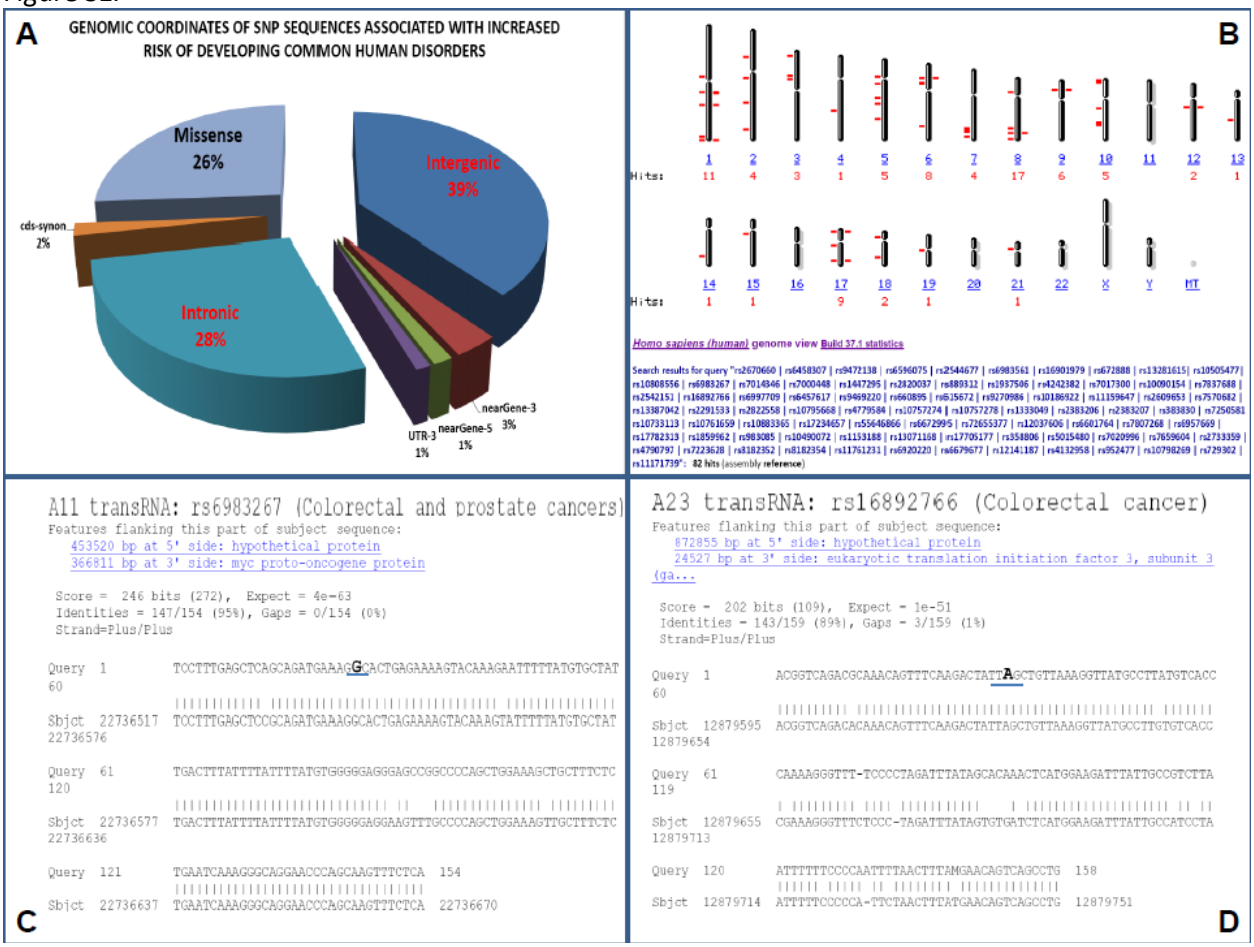
Bottom right figure shows hierarchical clustering of transRNA expression profiling data. Note that expression of selected transRNA-regulated “stemness” microRNAs induces expression patterns of cancer-susceptibility transRNAs in mesenchymal BJ1 cells which recapitulate transRNA expression profiles in epithelial cells, RWPE1 and HME1, and BJ1 cells engineered to stably express *NALP1*-locus transRNA A-allele (BJ1_2 cells).

F, Forced expression of selected transRNA-regulated microRNAs recapitulate transRNA-induced phenotypic changes in human cells. Note that the increased clonogenic growth potential which is conferred by stable expression of the ancestral A-allele compared to the pathology-linked G-allele (ref. 4 and two far left bars in the bottom panel) is recapitulated by stable expression of the A-allele-induced microRNAs, miR-205 and miR-20b. Average values of triplicate measurements of biological replicates normalized to control values obtained for BJ1/EGFP cells are shown. Top figures show typical colony staining patterns of one of the experiments.

Supplemental Figure S10. Identification of “stemness” microRNA signatures in blood-borne human prostate carcinoma (PC) metastasis precursor cells (Xp11; 4q21 clone in panels A; C), clinical PC samples (B; H), human PC cell lines selected for increased malignant potential in vivo (F; G), and normal human BJ cells engineered to constitutively express *NALP1*-locus autoimmunity transRNAs (Figure 1, main text). Genome-wide micro-RNA expression profiles for indicated cell lines, clinical samples, and human embryonic cell lines (hESC; panels A; B; H) were obtained using microarray, Q-RT-PCR, and high-throughput sequencing technologies and utilized to identify “stemness” micro-RNA signatures. Common sets of top-ranked up-regulated

“stemness” micro-RNAs (arrows) were identified for multiple pair-wise comparisons using hESC expression profiles as multidimensional reference vectors (1). Using this strategy, miR-20b, miR-375, miR-205, and miR-486 were selected for functional validation experiments (1). Panel D shows the results of luciferase reporter assay for assessment of miR-205 activity. Note that miR-205 activity is significantly higher in triple-transfected cells (miR-205; miR-375; miR-20b) compared to single-transfected cells (miR-205) despite having similar levels of miR-205 expression. Consistent with the proposed role in promoting malignant features of normal prostate epithelial cells, *NALP1*-locus SNP-RNA-regulated ‘stemness’ microRNAs enhance colony formation ability in agar of RWPE1 normal prostate epithelial cells (I).

Figure S1.


A
GENOMIC COORDINATES OF SNP SEQUENCES ASSOCIATED WITH INCREASED RISK OF DEVELOPING COMMON HUMAN DISORDERS
Missense 26%
Intergenic 39%
cds-synon 2%
Intronic 28%
nearGene-3 3%
nearGene-5 1%
UTR-3 1%
B
Hits:
1 2 3 4 5 6 7 8 9 10 11 12 13
11 4 3 1 5 8 4 17 6 5 2 1
14 15 16 17 18 19 20 21 22 X Y MT
1 1 9 2 1 1
Homo sapiens (human) genome view Build 37.1 statistics
C
A11 transRNA: rs6983267 (Colorectal and prostate cancers)
Features flanking this part of subject sequence:
453520 bp at 5' side: hypothetical protein
366811 bp at 3' side: myc proto-oncogene protein
Score = 246 bits (272), Expect = 4e-63
Identities = 147/154 (95%), Gaps = 0/154 (0%)
Strand=Plus/Plus
Query 1 TCCTTTGAGCTCAGCAGATGAAAGGCACTGAGAAAAGTACAAAGAATTTTTATGTGCTAT 60
Sbjct 22736517 TCCTTTGAGCTCCGCAGATGAAAGGCACTGAGAAAAGTACAAAGTATTTTTATGTGCTAT 22736576
Query 61 TGACTTTATTTTATTTTATGTGGGGGAGGGAGCCGGCCCCAGCTGGAAAGCTGCTTTCTC 120
Sbjct 22736577 TGACTTTATTTTATTTTATGTGGGGGAGGAAGTTTGCCCCAGCTGGAAAGTTGCTTTCTC 22736636
Query 121 TGAATCAAAGGGCAGGAACCCAGCAAGTTTCTCA 154
Sbjct 22736637 TGAATCAAAGGGCAGGAACCCAGCAAGTTTCTCA 22736670
D
A23 transRNA: rs16892766 (Colorectal cancer)
Features flanking this part of subject sequence:
872855 bp at 5' side: hypothetical protein
24527 bp at 3' side: eukaryotic translation initiation factor 3, subunit 3 (ga...
Score = 202 bits (109), Expect = 1e-51
Identities = 143/159 (89%), Gaps = 3/159 (1%)
Strand=Plus/Plus
Query 1 ACGGTCAGACGCAAACAGTTTCAAGACTATTAGCTGTTAAAGGTTATGCCTTATGTCACC 60
Sbjct 12879595 ACGGTCAGACACAAACAGTTTCAAGACTATTAGCTGTTAAAGGTTATGCCTTGTGTCACC 12879654
Query 61 CAAAAGGGTTT-TCCCCTAGATTTATAGCACAAACTCATGGAAGATTTATTGCCGTCTTA 119
Sbjct 12879655 CGAAAGGGTTTCTCCC-TAGATTTATAGTGTGATCTCATGGAAGATTTATTGCCATCCTA 12879713
Query 120 ATTTTTTCCCCAATTTTAACTTTAMGAACAGTCAGCCTG 158
Sbjct 12879714 ATTTTTCCCCCA-TTCTAACTTTATGAACAGTCAGCCTG 12879751

NALP1 locus transRNA (rs2670660)

Features flanking this part of subject sequence:
85717 bp at 5' side: NACHT-, LRR-, and PYD-containing 1 paralog c
392858 bp at 3' side: WSC domain containing 1

Score = 50.0 bits (54), Expect = 6e-04
Identities = 53/70 (75%), Gaps = 0/70 (0%)
Strand=Plus/Plus

```
Query  1         CCACGCACAAGTGATCTACCAGTCTTTTAAAATTCTAATTATTAAAACCCAAACATGCTCT  60
                 ||| || ||||  | ||||   ||||||||||||||| ||  |||| | | ||||||| |
Sbjct  36685302  CCATGCTCAAGAAACCTACTGATCTTTTAAAATTCTGTTTCTAAATCTCCAACATGCTTT  36685361

Query  61        TTCATTTCCA  70
                  ||||||||
Sbjct  36685362  AGCATTTCCA  36685371
```

E3 transRNA: rs3803662 (Breast cancer)

Features flanking this part of subject sequence:
5742 bp at 5' side: trinucleotide repeat containing 9
382864 bp at 3' side: hypothetical protein

Score = 134 bits (148), Expect = 3e-29
Identities = 136/174 (78%), Gaps = 10/174 (5%)
Strand=Plus/Minus

```
Query  52        GATCTGTCAT-AGAAGGGTTTAATTATATCTGCCTAATGATTTTCTCTCCTTAATGCCTC  110
                 |||||||||| |||||| || |  |||||  |  ||| ||||||||||||||| || || |
Sbjct  17711113  GATCTGTCATGAGAAGGCTTCATGTATATTGGTTTAAGGATTTTCTCTCCTTCATTCCCC  17711054

Query  111       TATAGCTGTCCCTTAGCGAAGAATAAAACTGTGGACTGACC-----CCCACCCATTTGCG  165
                 ||||||| | |||||| ||| ||||||| | |||| |||||     ||||||||||
Sbjct  17711053  TATAGCTATTCCTTAGTGAAAAATAAAA-TCTGGATTGACCCCCCTCCCACCCATTTTGA  17710995

Query  166       AAGAAAGTACTGGGTCTTCAGCTTTCATTGTTCAGCCGGTGGTCTTTGTGGACA  219
                  ||||| ||||| |||||||| |||||||| |||   ||| ||||||| ||
Sbjct  17710994  GAGAAAATACTGTGTCTTCAGCCTTCATTGTTTAGC---TGGGCTTTGTGGGCA  17710944
```

E12 transRNA: rs2736990 (Parkinson's disease)

Features in this part of subject sequence:
synuclein, alpha
synuclein, alpha

Score = 75.2 bits (82), Expect = 2e-11
Identities = 102/142 (71%), Gaps = 9/142 (6%)
Strand=Plus/Minus

```
Query  1         ATGTCTGCCTTGCATCAGATAATGGCTTACAAGTTAATCTCCTCTTG-CTCCCTGTTAC  59
                 |||| ||| |||||||| ||||||| |||||  |||| |  || || ||||  ||
Sbjct  13024796  ATGTTTGCTTTTGCATCAGGTAATGGCATACAACATAATTTGTTCCTGACTCCCCACTAT  13024737

Query  60        ACACATATACATCTTCTTCCTAAACAGCTCATAAGGTGAAAGAAAGACTCAGATTTCTGA  119
                 ||||| ||| |||| |||    |  |||| ||||  ||| ||| ||||   ||||||||
Sbjct  13024736  ACACACATATATCTCCTTTGACATTAGCTAATAAAATGACAGAGAGACGTTGATTTCTGA  13024677

Query  120       CTATGTAAT7GATAATATCACA  141
                  |       ||||||||||||
Sbjct  13024676  C--------TGATAATATCACA  13024663
```

A16 transRNA: rs889312 (Breast cancer)

Features flanking this part of subject sequence:
59455 bp at 5' side: mitogen activated protein kinase kinase kinase 1
345311 bp at 3' side: hypothetical protein

Score = 89.7 bits (98), Expect = 2e-16
Identities = 60/67 (89%), Gaps = 0/67 (0%)
Strand=Plus/Minus

```
Query  1         ATGCCCCTGCTGGAGAAAGGAATGTGCAAATTAAGAGACTACAAATCAGTTTGAAAACTC  60
                 |||||||||||||||||||||||||| |||| |||| |||||||||| | ||||||||
Sbjct  19867347  ATGCCCCTGCTGGAGAAAGGAATGTGTAAATCAAGACACTACAAATCAATGTGAAAACTC  19867288

Query  61        AACGACT  67
                  |  ||||
Sbjct  19867287  ACTGACT  19867281
```

A9 transRNA: rs10505477 (Prostate cancer)

Features flanking this part of subject sequence:
446299 bp at 5' side: hypothetical protein
374115 bp at 3' side: myc proto-oncogene protein

Score = 57.2 bits (62), Expect = 3e-06
Identities = 55/71 (77%), Gaps = 0/71 (0%)
Strand=Plus/Minus

```
Query  37        TGTCATGTGTCACCACTTGTCTATCAAACAGGAAGCCTTAATTGGAGATGAAGATTTAGA  96
                 |||| |  |||| || ||  |||| ||||||||||||||  |||  |||| |||||||||
Sbjct  22729366  TGTCTTCAGTCAGCATTTACCTATTAAACAGGAAGCCTTCTTTGAGGATGGAGATTTAGA  22729307

Query  97        AAAGGGGCAAA  107
                  |||  | ||||
Sbjct  22729306  AAAAAGACAAA  22729296
```

A6 transRNA: rs16901979 (Prostate cancer)

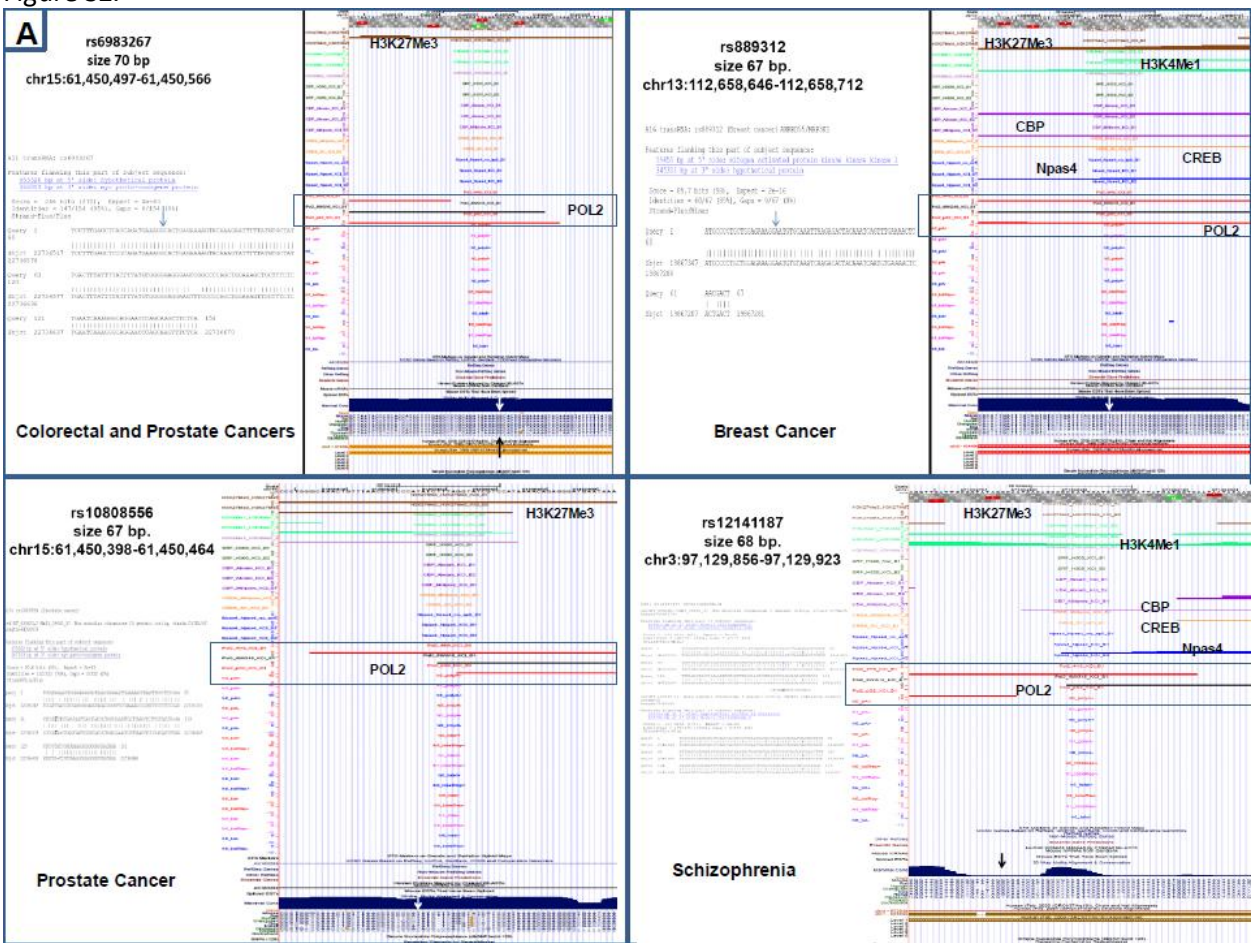
Features flanking this part of subject sequence:
190292 bp at 5' side: hypothetical protein
630042 bp at 3' side: myc proto-oncogene protein

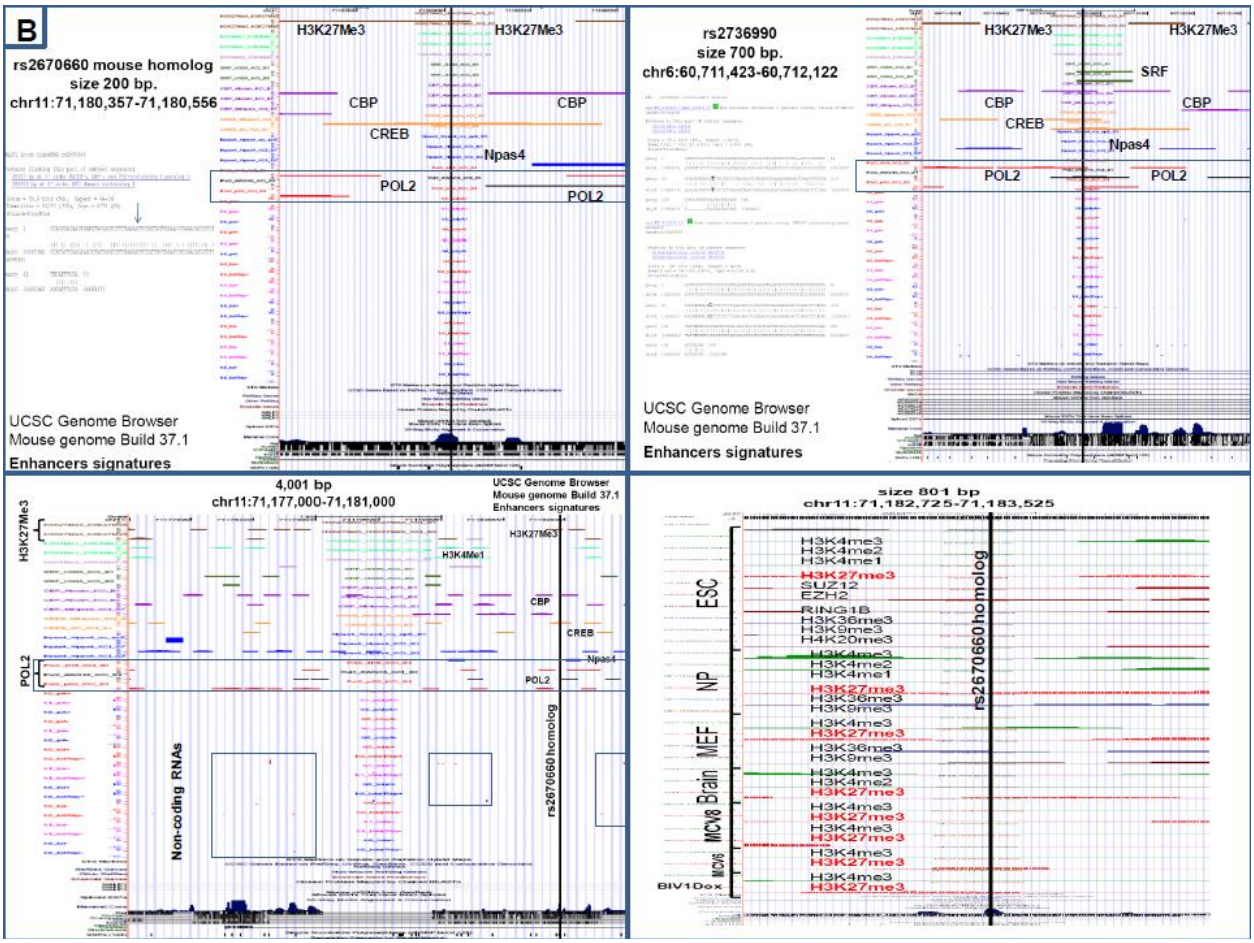
Score = 82.4 bits (90), Expect = 1e-13
Identities = 111/152 (73%), Gaps = 2/152 (1%)
Strand=Plus/Plus

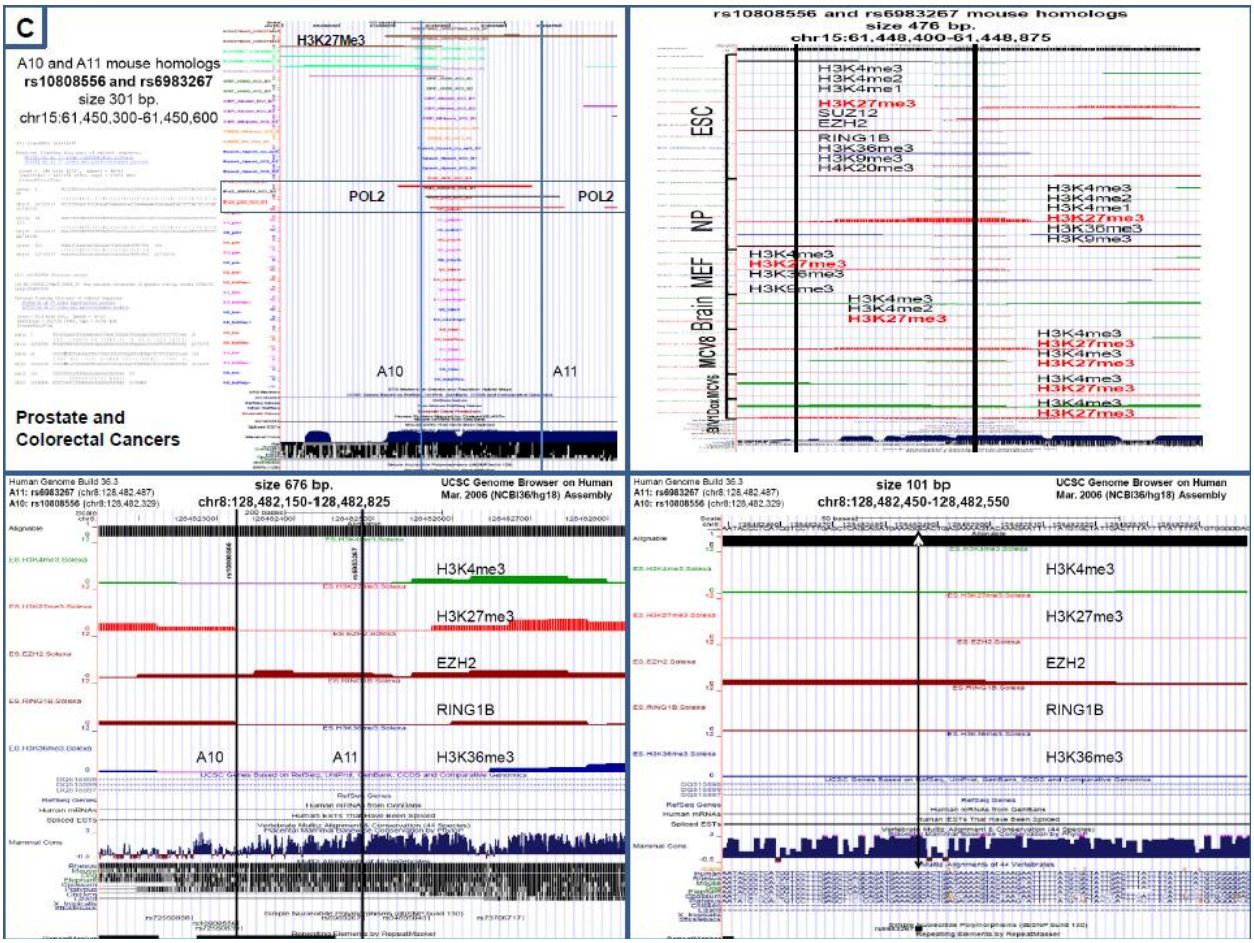
```
Query  36        AGCATTACTTATATCTGGCAAATGGTATTTTTGAGATAACATGTTATGGAAGAAAGTGAA  95
                 |||||||||||||  ||||||   || || | |||||   || | ||| |  |||||||
Sbjct  22473289  AGCATTACTTATATAGGGCAAACACTAGTTAT-AGATATTGTGATGTGGTAAGAAGTGAA  22473347

Query  96        CTGAACTTGGAAGTTTGAAGATCTCGATTGAAGTATCATTTCTG-CCTCAACTACTTGCA  154
                 |   ||||||||  |||||||| | ||| |||| ||  |||| |||   || ||||
Sbjct  22473348  CCAGACTTGGAAAAATTGAAGATCTGGGTTGGAGTACCAGGTCTGCCCTTCCCTGCTTGTT  22473407

Query  155       TTAACTTGTACAAGTCATTCAACCGCTCTGAA  186
                 | |||||||| | ||| |||| | | |||||
Sbjct  22473408  TGAACTTGTACCATTCACTCAAGCACCCTGAA  22473439
```

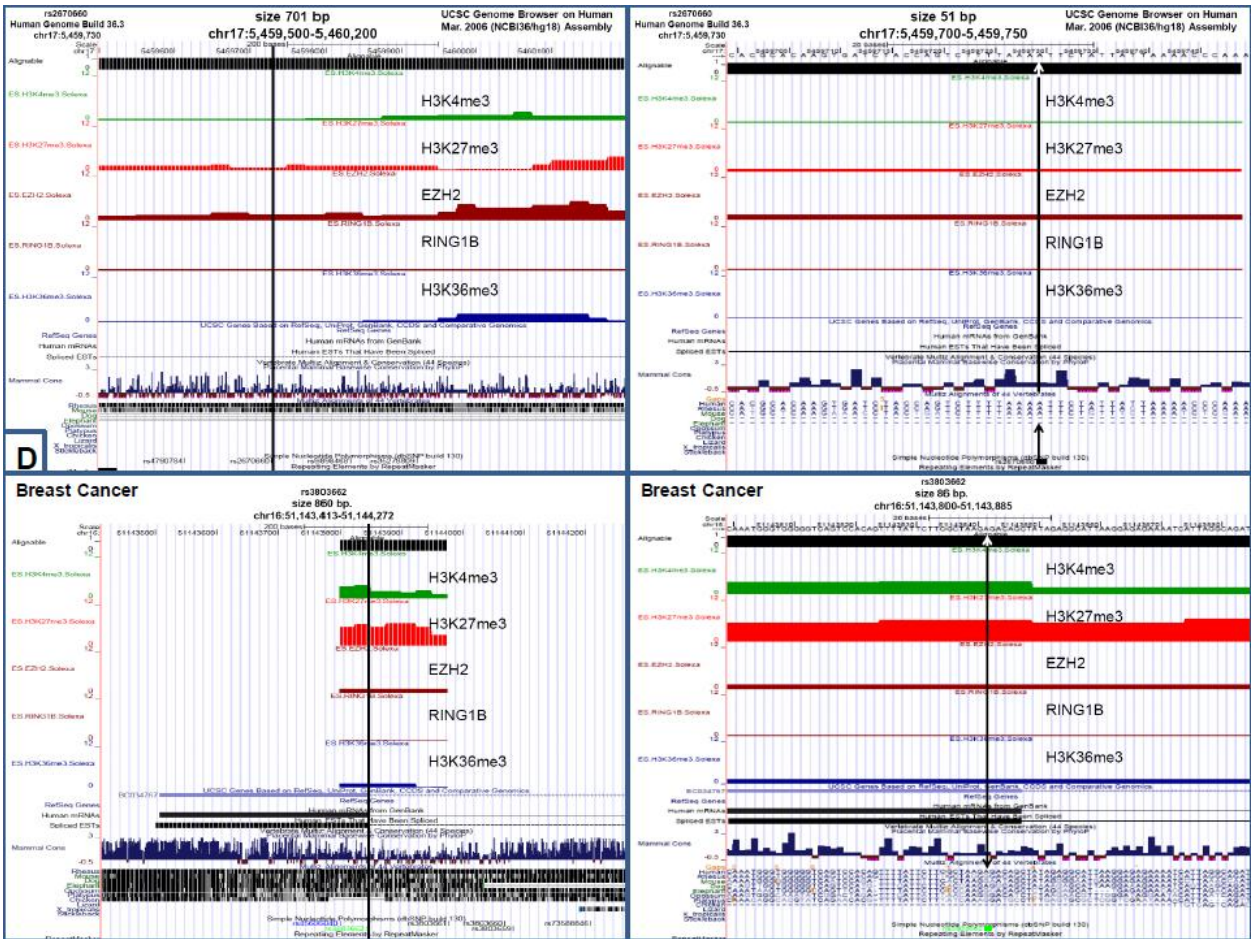

A10 transRNA: rs10808556 (Prostate cancer)
ref|NT_039621.7|Mm15_39661_37 Mus musculus chromosome 15 genomic contig, strain C57BL/6J
Length=64780978
Features flanking this part of subject sequence:
453322 bp at 5' side: hypothetical protein
367013 bp at 3' side: myc proto-oncogene protein
Score = 80.6 bits (88), Expect = 3e-13
Identities = 112/152 (73%), Gaps = 4/152 (2%)
Strand=Plus/Plus
Query 2 TCCATAGAGCCTGCAGAGGGCACTAGACTGGGAATTAGAAAACCTGATTTCCCTTCC-AG 60
Sbjct 22736319 TCCATTTATCCTGCAGGGGGAGCTAGACCGGGTTCTGAAAACCCCGGTTTCTCTCCGAG 22736378
Query 61 CTCCATCTCTGACCAATTGCCTGACCCTGGTCAAATTGCTTAACCTCTTCCTATCTC-AG 119
Sbjct 22736379 CCCCATA-CTGGGTGATTGTGTGACCCTGGGCAAACTGTTTAACCTCCCCATATCTTTAG 22736437
Query 120 CTCCCTATCCATAAAACAGAGGGACGAATAAA 151
Sbjct 22736438 GTATCG-TCCATAAAACAGAGGGATGAATAAA 22736468
B6 transRNA: rs2291533 (Breast cancer)
ref|NT_039595.7|Mm14_39635_37 Mus musculus chromosome 14 genomic contig, strain C57BL/6J
Length=17238926
Features flanking this part of subject sequence:
33894 bp at 5' side: hypothetical protein
811 bp at 3' side: Ngg1 interacting factor 3 like 1 binding protein 1 isoform 1
Score = 129 bits (142), Expect = 6e-28
Identities = 98/116 (84%), Gaps = 0/116 (0%)
Strand=Plus/Minus
Query 10 TTATACTTCCTCATGGTTCTCTTGGATATCCTCTGGAACTGTTTAGAAGACTGAAGAATT 69
Sbjct 11780963 TTATACTTCCTCGTGGTTCTCTTGGAAAGCCTTTGGAATTGCTTCGAAGACTGAAGGATC 11780904
Query 70 TCATCCCCCAGAAACTCACACTGTTGAAGCTCAGCATGTCTTTGGGCCAGTAGCTT 125
Sbjct 11780903 TCAGCCCCTAGAAACTCACACTGCTTAAGCTCAGCGTGCCTTTGTTCCAGAAGCTT 11780848
D26 transRNA: rs12141187 (Schizophrenia)
ref|NT_039240.7|Mm3_39280_37 Mus musculus chromosome 3 genomic contig, strain C57BL/6J
Length=80458789
Features flanking this part of subject sequence:
109512 bp at 5' side: B-cell CLL/lymphoma 9
84165 bp at 3' side: olfactory receptor 1402
Score = 169 bits (91), Expect = 5e-40
Identities = 149/177 (84%), Gaps = 4/177 (2%)
Strand=Plus/Minus
Query 1 TCAGCATCAGTCACCTCAGCCAGGTCC-CTGAATCACAGCCAAGCCTAGA7GAGTGGTAT 59
Sbjct 46512328 TCAGCATCCGTCACTTCAGCCAGG-CGTGTGAATCACAGCCAAATCTAGATGACTTGTAT 46512270
Query 60 TATTGACCATGATAATGGGAGGATGAATGGTGGCTATGACTGCCTG--CTGCAATCAACC 117
Sbjct 46512269 TATTGACCATGATAATGGGAGGATGAATGGCGGCTGTGATTGCCTATACCGCAATCAACC 46512210
Query 118 TTTAGGATGGCCAGAAATTCTGATTTGGCCAGCCCTTGGCCCAGACAGCAATGTCCC 174
Sbjct 46512209 TTTAAGATGGTTAGGAATCCTGATCCGGCCAGTCCGTGGCCCAGATAGTGATGTCCC 46512153
D2 transRNA: rs1153188 (Type 2 diabetes)
ref|NT_166468.1|MmUn_163642_37 Mus musculus genomic contig, strain C57BL/6J
Length=50955
Score = 71.6 bits (78), Expect = 2e-10
Identities = 99/133 (74%), Gaps = 7/133 (5%)
Strand=Plus/Plus
Query 2 AAGATGGTCTGAATGGCAAAATGGATAAAATTAAAATCAAAACTAGTGAACTGAAATAGC 61
Sbjct 31409 AAGATGGTCTGGATG-CAAAATGAATAAAATTAAAGACCAAAATAGTGAGCTGAAGCAGA 31467
Query 62 AAGGTGAGA-AGTTCTTCTGAAATGCAGTATaaaagataaaaagaaatacaaagaaaaag 120
Sbjct 31468 AAATTGAGAGACTTC-CCTCATCTTCAGCATAAAAG---AGAAGAAGTTCGCAGAGAAGG 31523
Query 121 TCATG-AAGGACA 132
Sbjct 31524 ACATGAAAGGACA 31536
D5 transRNA: rs17705177 (Type 2 diabetes)
ref|NT_039624.7|Mm16_39664_37 Mus musculus chromosome 16 genomic contig, strain C57BL/6J
Length=62254923
Features flanking this part of subject sequence:
130191 bp at 5' side: B-cell leukemia/lymphoma 6
50370 bp at 3' side: hypothetical protein
Score = 50.0 bits (54), Expect = 7e-04
Identities = 42/52 (80%), Gaps = 0/52 (0%)
Strand=Plus/Plus
Query 35 AGGGTTATTGTGAAGATTAAAGTGGAATGTGCATGCAAAAGTACTTTGCAAA 86
Sbjct 21107952 AGGGTTATTGTGAAGATTAAAGAGGATAATGGACGTGAAAGCACTTTGGAAA 21108003
C12 transRNA: rs6957669 (Multiple sclerosis)
ref|NT_039578.7|Mm13_39618_37 Mus musculus chromosome 13 genomic contig, strain C57BL/6J
Length=28633595
Features in this part of subject sequence: Uqcrfs1
ubiquinol-cytochrome c reductase, Rieske iron-sulfur poly...
ref|NM_025710.1| Mus musculus ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1 (Uqcrfs1), mRNA
Length=1316
Score = 100 bits (110), Expect = 4e-19
Identities = 124/169 (73%), Gaps = 2/169 (1%)
Strand=Plus/Plus
Query 1 TGGTGGTGATTACTGCCCTTGCTGGGGGTCACACA--GATGCATCTGGGAGGATCTGGAA 58
Sbjct 765 TGGTGGCTATTATTGCCCCTGCCATGGGTCACACTATGAAGCCTCTGGCAGGATCAGGAA 824
Query 59 GGGGCCTGCCCCTCTTGAGCTTGGAGCTCCCTCATATGAGTTCACCAGTGAGGACACAGT 118
Sbjct 825 GGGCCCTGCACCTCTCAACCTGGAAGTGCCTGCCTACGAGTTCACCAGTGATGATGTAGT 884
Query 119 CATTGTTGGTTAGAGACTGGGACTCAAGTTGTAGGCTCCTTTCAGTCTT 167
Sbjct 885 TGTTGTGGGTTAGACACGTGGACTCAGGGCCTAGGTTTCTGTCGGTCTT 933

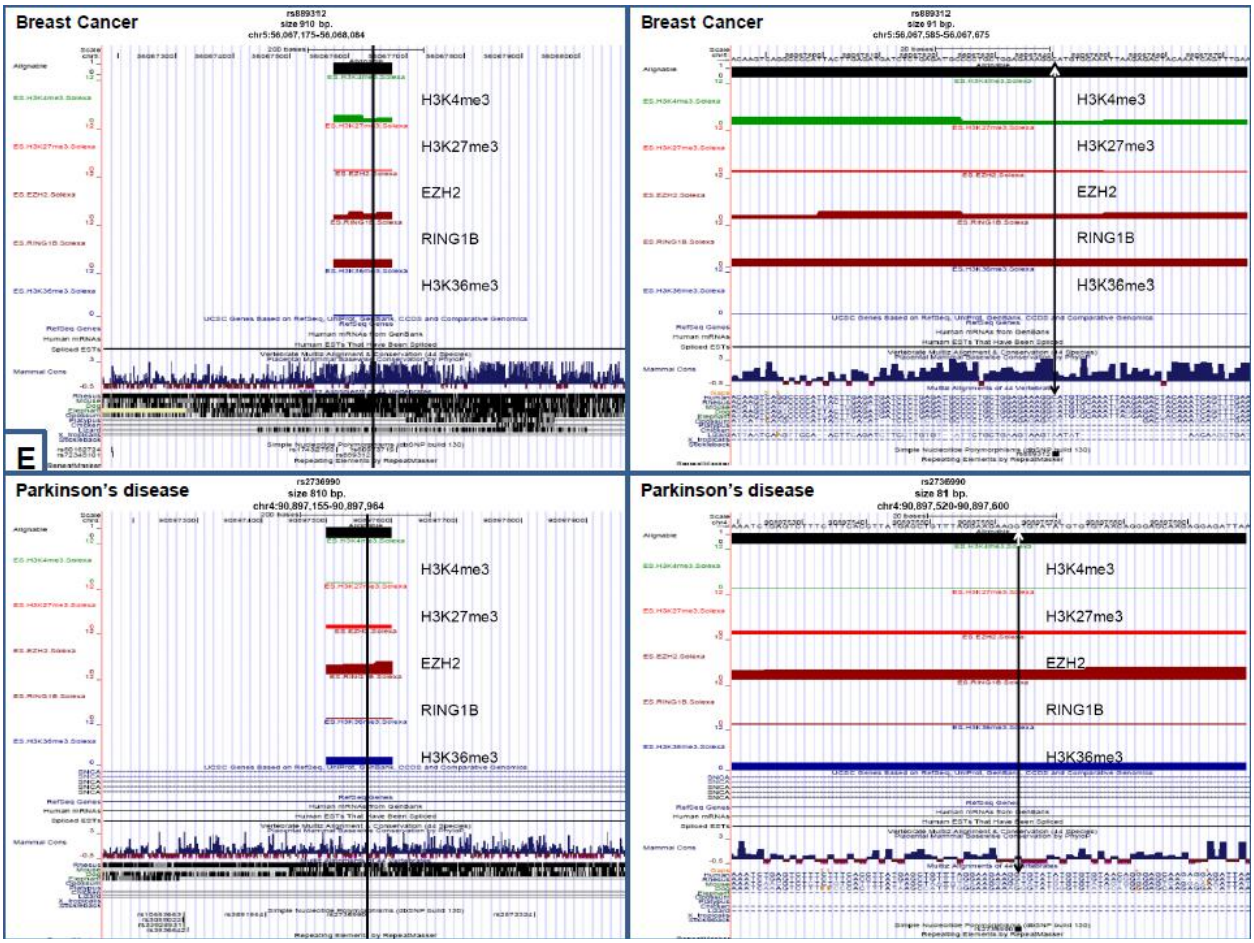
Figure S2.


A
rs6983267
size 70 bp
chr15:61,450,497-61,450,566
H3K27Me3
POL2
Colorectal and Prostate Cancers
rs889312
size 67 bp.
chr13:112,658,646-112,658,712
H3K27Me3
H3K4Me1
CBP
CREB
Npas4
POL2
Breast Cancer
rs10808556
size 67 bp.
chr15:61,450,398-61,450,464
H3K27Me3
POL2
Prostate Cancer
rs12141187
size 68 bp.
chr3:97,129,856-97,129,923
H3K27Me3
H3K4Me1
CBP
CREB
Npas4
POL2
Schizophrenia


B
rs2670660 mouse homolog
size 200 bp.
chr11:71,180,357-71,180,556
H3K27Me3
H3K27Me3
CBP
CBP
CREB
Npas4
POL2
POL2
UCSC Genome Browser
Mouse genome Build 37.1
Enhancers signatures
rs2736990
size 700 bp.
chr6:60,711,423-60,712,122
H3K27Me3
H3K27Me3
SRF
CBP
CBP
CREB
Npas4
POL2
POL2
UCSC Genome Browser
Mouse genome Build 37.1
Enhancers signatures
4,001 bp
chr11:71,177,000-71,181,000
UCSC Genome Browser
Mouse genome Build 37.1
Enhancers signatures
H3K27Me3
H3K4Me1
H3K27Me3
CBP
CREB
Npas4
POL2
POL2
Non-coding RNAs
rs2670660 homolog
size 801 bp
chr11:71,182,725-71,183,525
ESC
NP
MEF
Brain
MCv8
MCv5
BIV1Dox
H3K4me3
H3K4me2
H3K4me1
H3K27me3
SUZ12
EZH2
RING1B
H3K36me3
H3K9me3
H4K20me3
rs2670660 homolog


C
A10 and A11 mouse homologs
rs10808556 and rs6983267
size 301 bp.
chr15:61,450,300-61,450,600
H3K27Me3
POL2
POL2
A10
A11
Prostate and
Colorectal Cancers
rs10808556 and rs6983267 mouse homologs
size 476 bp.
chr15:61,448,400-61,448,875
ESC
NP
MEF
Brain
MCV8
H3K4me3
H3K4me2
H3K4me1
H3K27me3
SUZ12
EZH2
RING1B
H3K36me3
H3K9me3
H4K20me3
Human Genome Build 36.3
A11: rs6983267 (chr8:128,482,487)
A10: rs10808556 (chr8:128,482,329)
size 676 bp.
chr8:128,482,150-128,482,825
UCSC Genome Browser on Human
Mar. 2006 (NCBI36/hg18) Assembly
H3K4me3
H3K27me3
EZH2
RING1B
A10
A11
H3K36me3
Human Genome Build 36.3
A11: rs6983267 (chr8:128,482,487)
A10: rs10808556 (chr8:128,482,329)
size 101 bp
chr8:128,482,450-128,482,550
UCSC Genome Browser on Human
Mar. 2006 (NCBI36/hg18) Assembly
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3


rs2670660
Human Genome Build 36.3
chr17:5,459,730
size 701 bp
chr17:5,459,500-5,460,200
UCSC Genome Browser on Human
Mar. 2006 (NCBI36/hg18) Assembly
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
D
rs2670660
Human Genome Build 36.3
chr17:5,459,730
size 51 bp
chr17:5,459,700-5,459,750
UCSC Genome Browser on Human
Mar. 2006 (NCBI36/hg18) Assembly
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
Breast Cancer
rs3803662
size 860 bp.
chr16:51,143,413-51,144,272
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
Breast Cancer
rs3803662
size 86 bp.
chr16:51,143,800-51,143,885
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3


Breast Cancer
rs889312
size 910 bp.
chr5:56,067,175-56,068,084
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
E
Breast Cancer
rs889312
size 91 bp.
chr5:56,067,583-56,067,675
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
Parkinson's disease
rs2736990
size 810 bp.
chr4:90,897,155-90,897,964
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
Parkinson's disease
rs2736990
size 81 bp.
chr4:90,897,520-90,897,600
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3

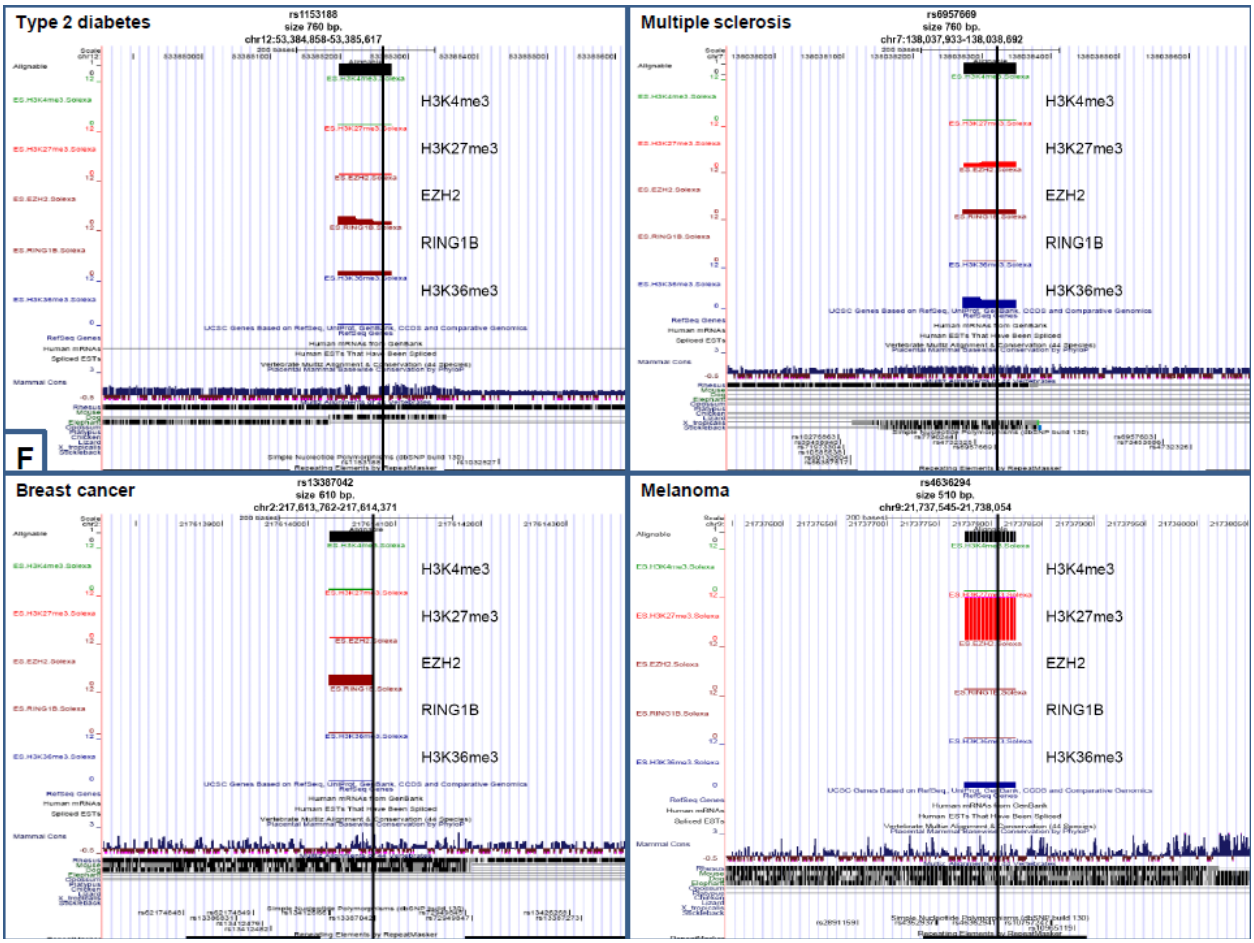
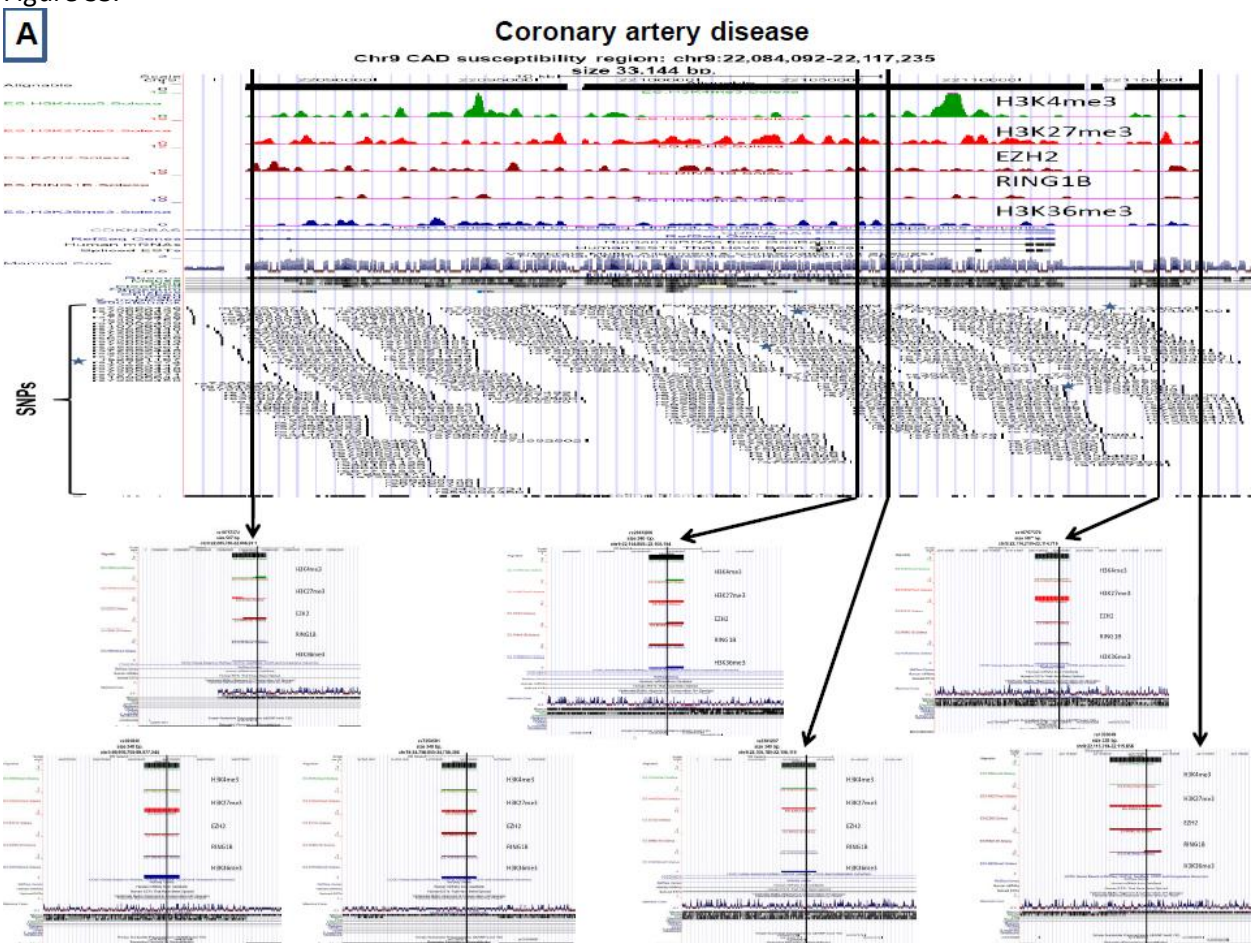
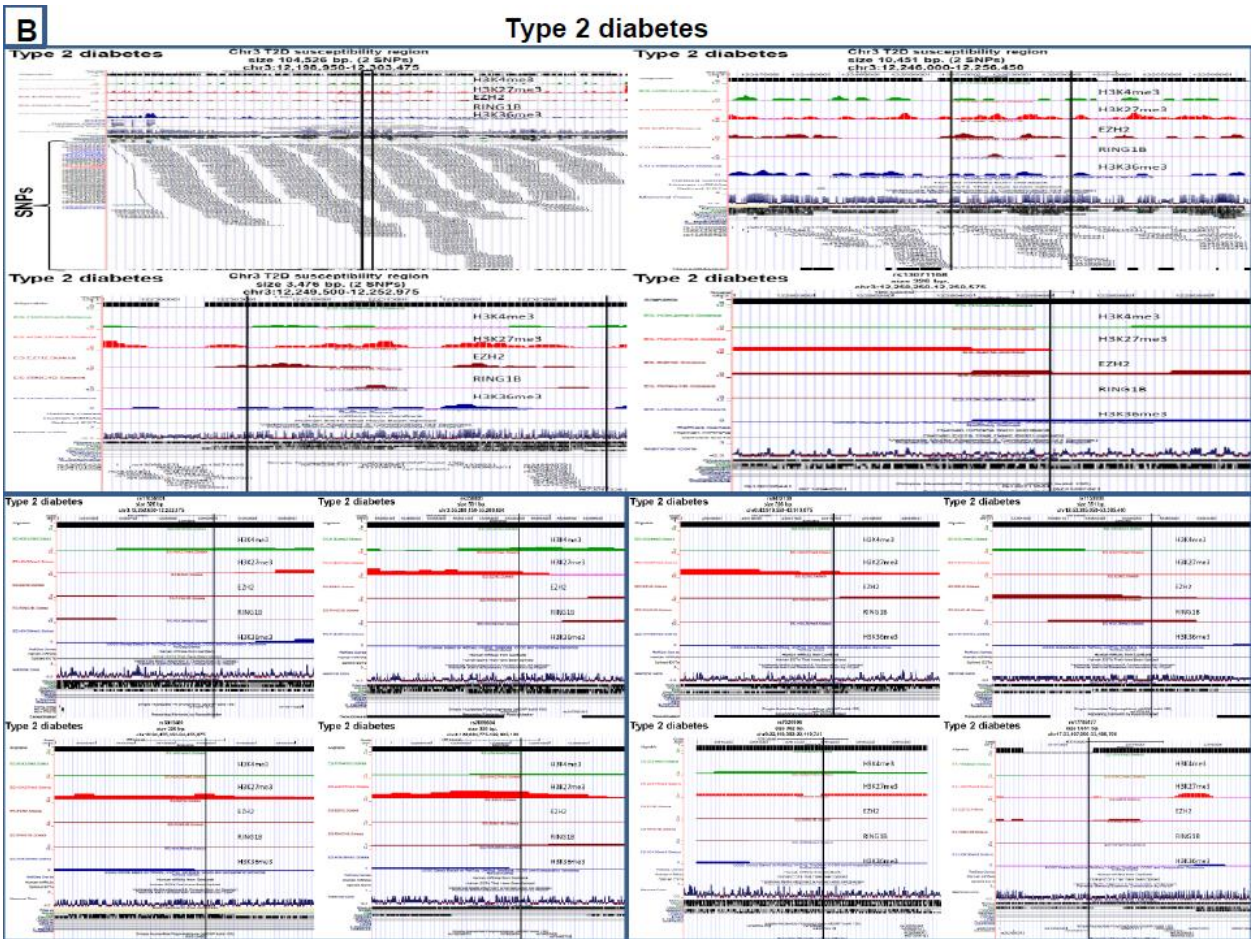

Type 2 diabetes
rs1153188
size 760 bp.
chr12:53,384,858-53,385,617
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
Multiple sclerosis
rs6957669
size 760 bp.
chr7:138,037,933-138,038,692
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
F
Breast cancer
rs13387042
size 610 bp.
chr2:217,613,762-217,614,371
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
Melanoma
rs4636294
size 510 bp.
chr9:21,737,545-21,738,054
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3

Figure S3.

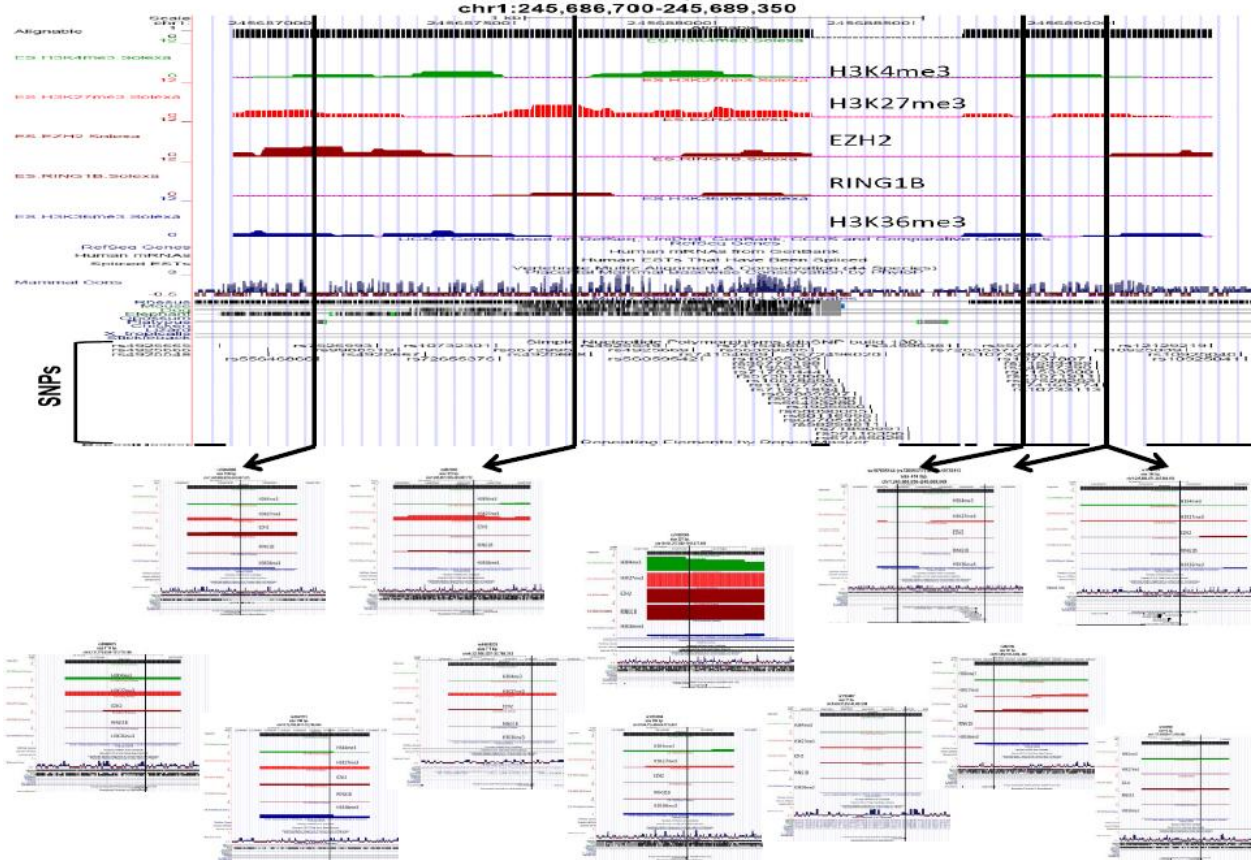

A
Coronary artery disease
Chr9 CAD susceptibility region: chr9:22,084,092-22,117,235
size 33,144 bp.
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
SNPs

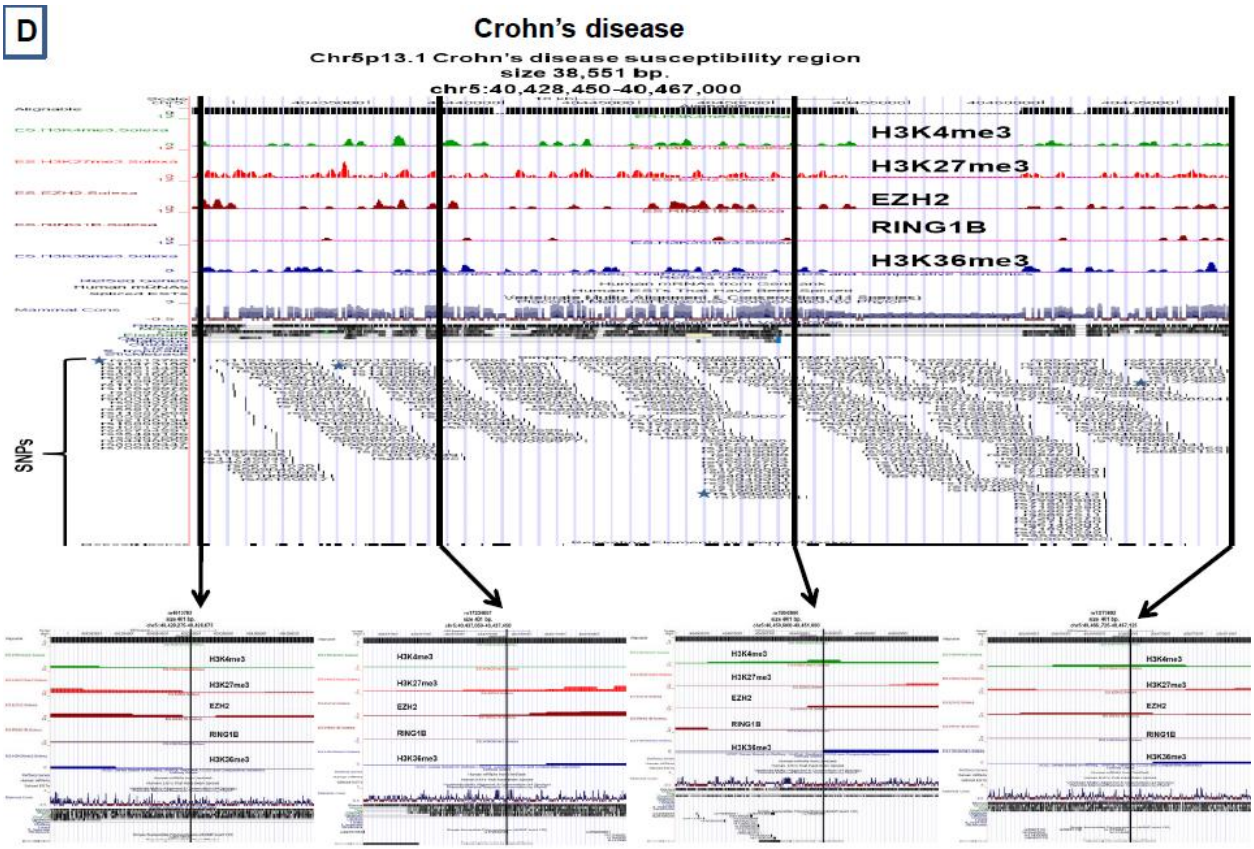

B
Type 2 diabetes
Type 2 diabetes
Chr3 T2D susceptibility region
size 104,526 bp. (2 SNPs)
chr3:12,198,950-12,303,475
SNPs
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
Type 2 diabetes
Chr3 T2D susceptibility region
size 10,451 bp. (2 SNPs)
chr3:12,246,000-12,256,450
Type 2 diabetes
Chr3 T2D susceptibility region
size 3,478 bp. (2 SNPs)
chr3:12,249,500-12,252,975
Type 2 diabetes
Type 2 diabetes
Type 2 diabetes
Type 2 diabetes
Type 2 diabetes
Type 2 diabetes
Type 2 diabetes
Type 2 diabetes

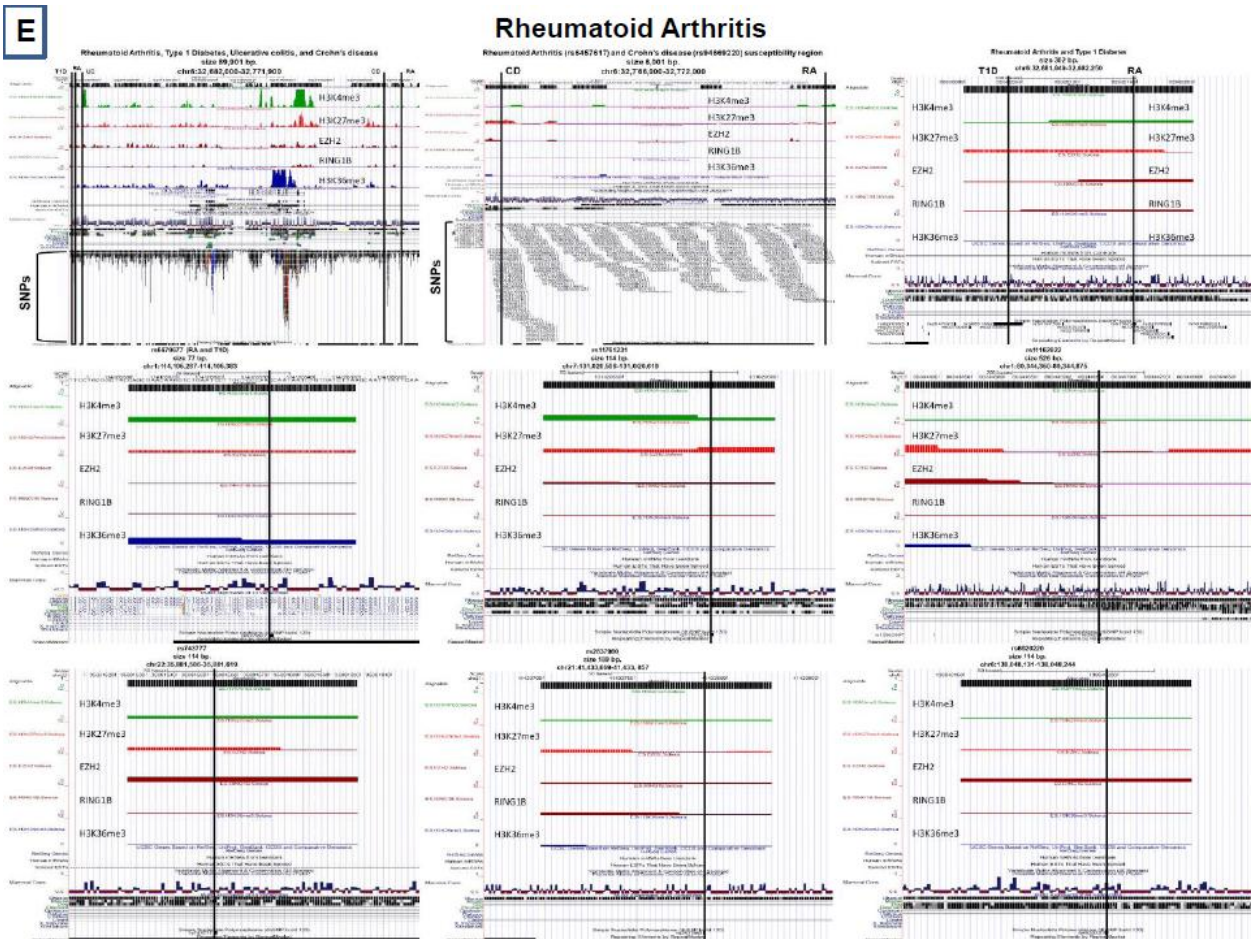
C

Crohn's disease

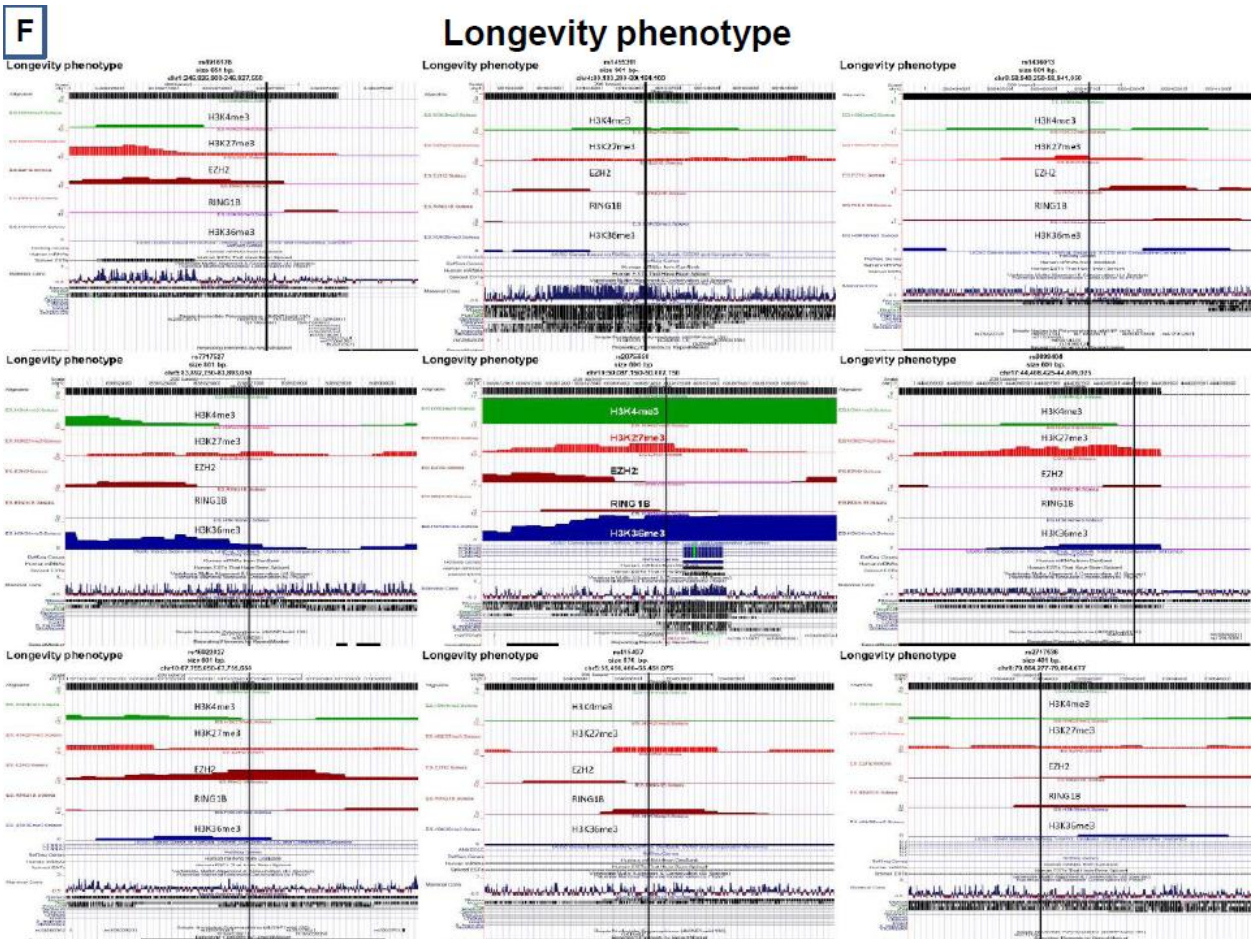
Crohn's disease susceptibility region
size 2,651 bp.
chr1:245,686,700-245,689,350




D
Crohn's disease
Chr5p13.1 Crohn's disease susceptibility region
size 38,551 bp.
chr5:40,428,450-40,467,000
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
SNPs


E
Rheumatoid Arthritis
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
SNPs
CD
RA
T1D

F


Longevity phenotype
Longevity phenotype
Longevity phenotype
Longevity phenotype
Longevity phenotype
Longevity phenotype
Longevity phenotype
Longevity phenotype
Longevity phenotype
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3

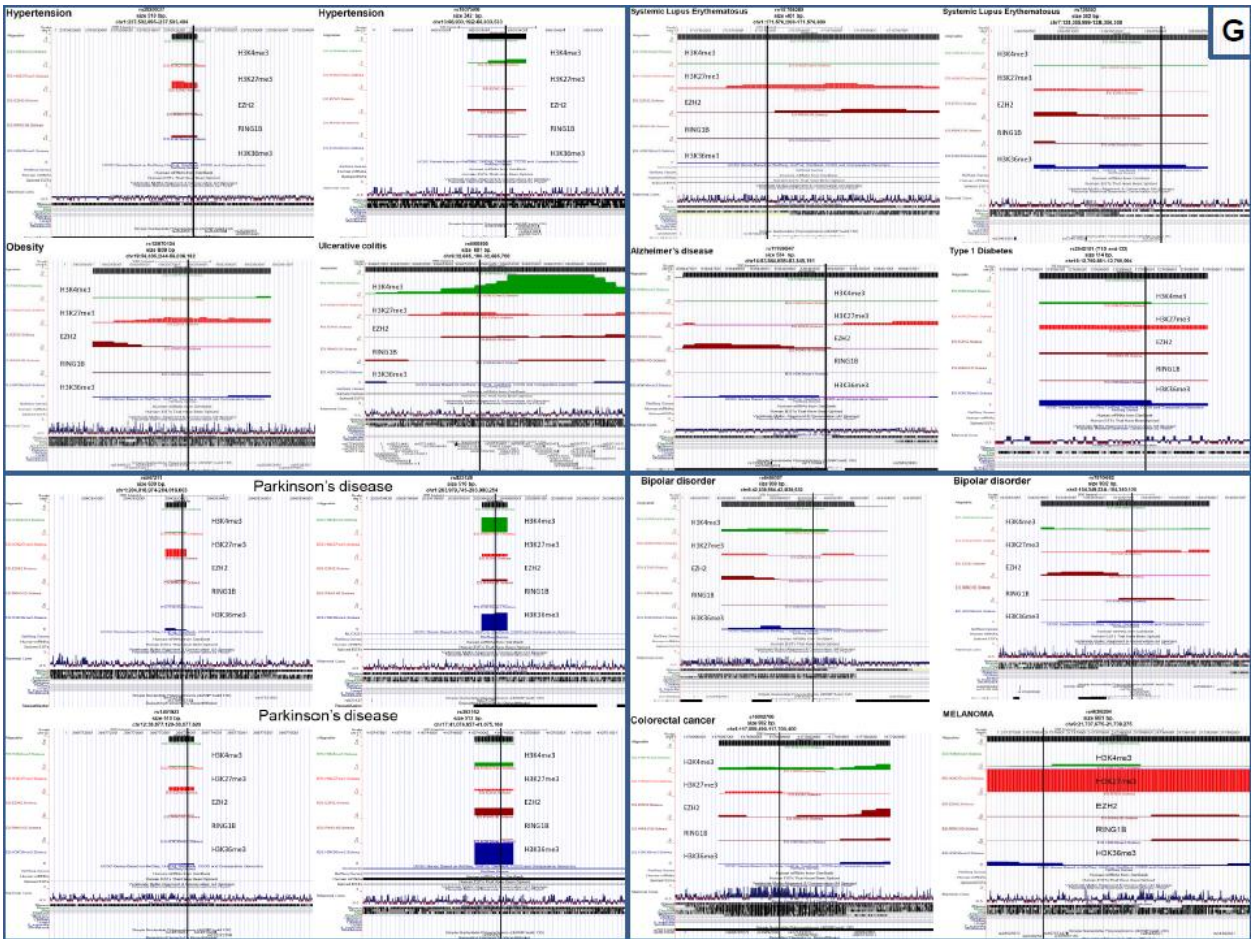
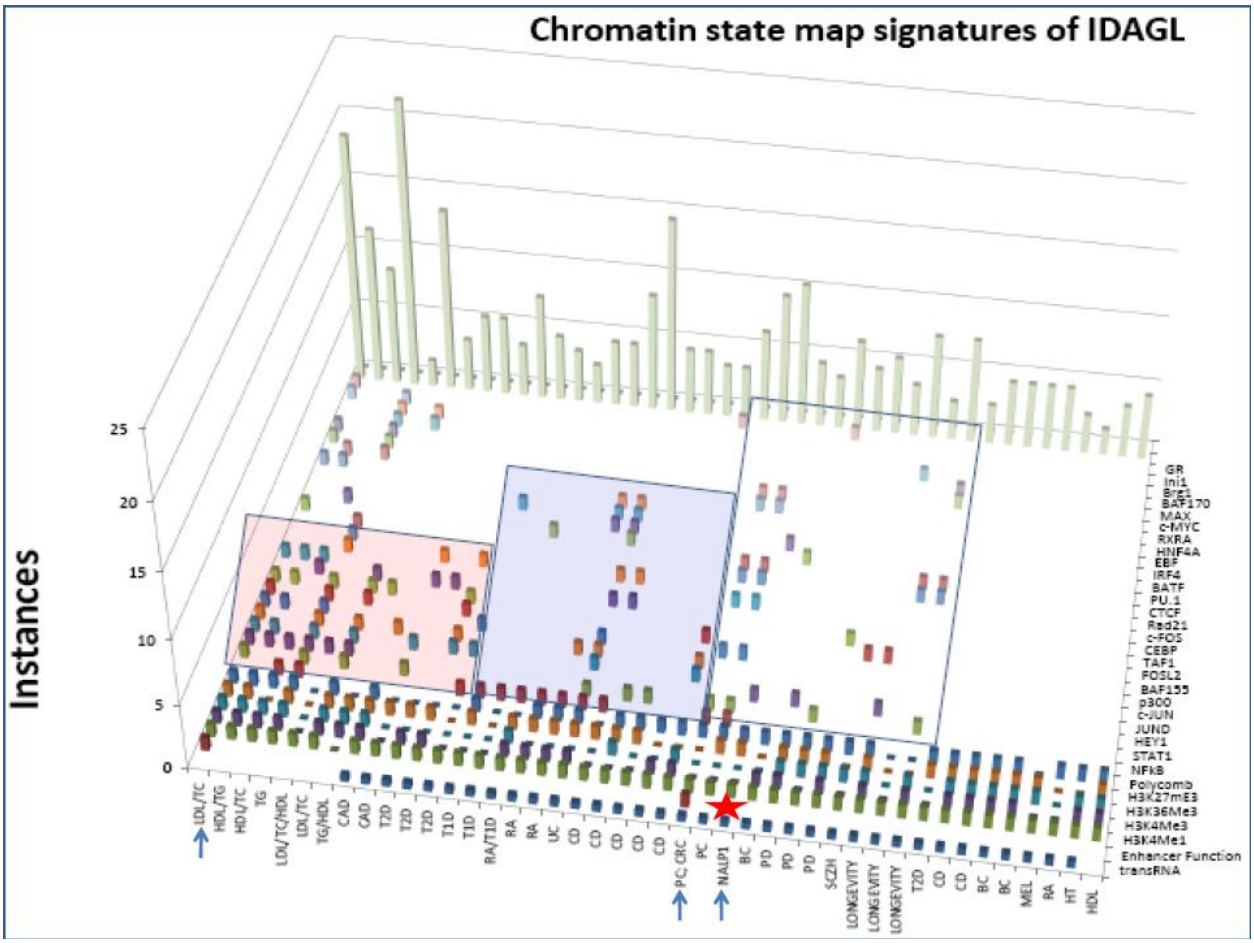
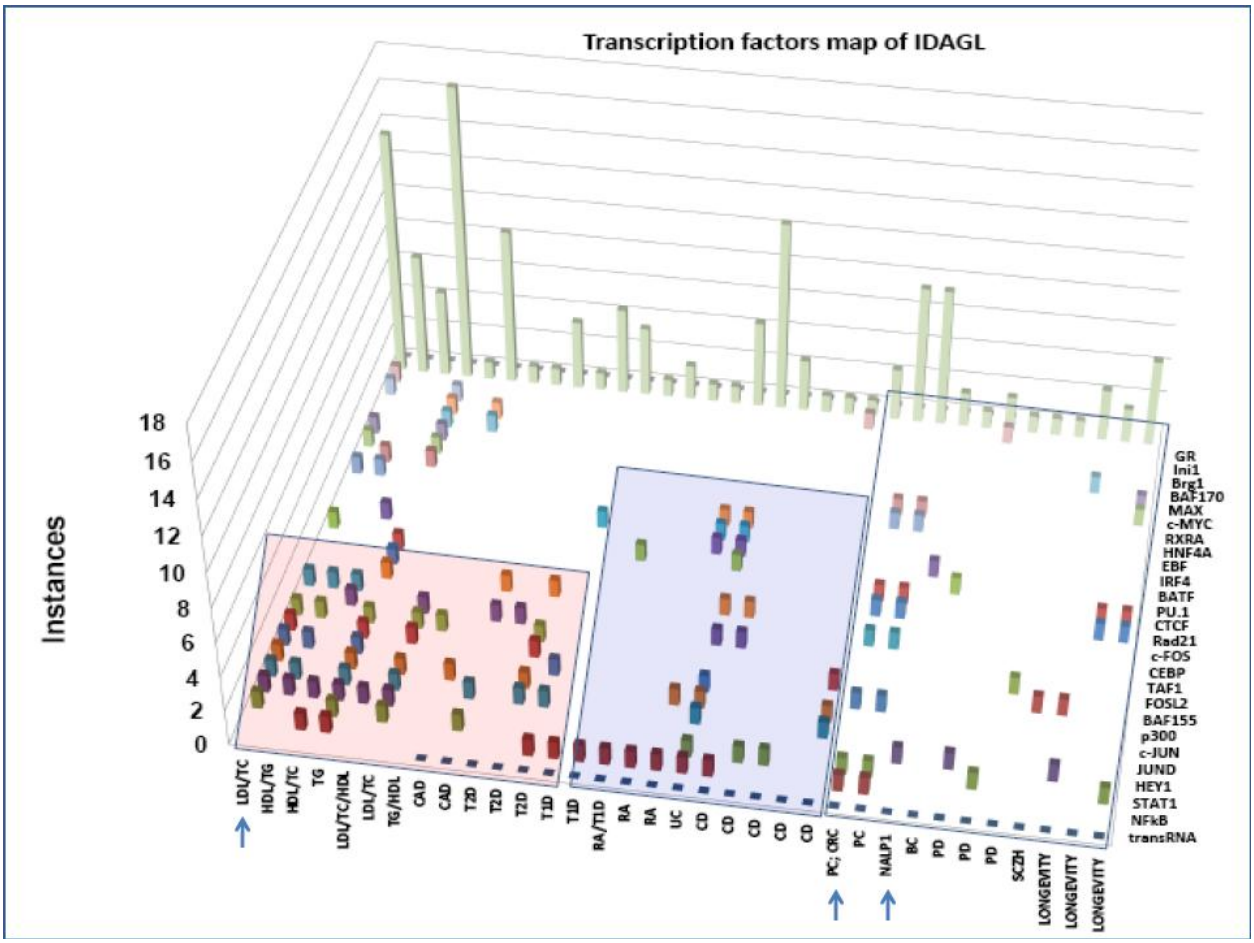

G
Hypertension
Hypertension
Systemic Lupus Erythematosus
Systemic Lupus Erythematosus
Obesity
Ulcerative colitis
Alzheimer's disease
Type 1 Diabetes
Parkinson's disease
Parkinson's disease
Bipolar disorder
Bipolar disorder
Colorectal cancer
MELANOMA
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3

Figure S4.

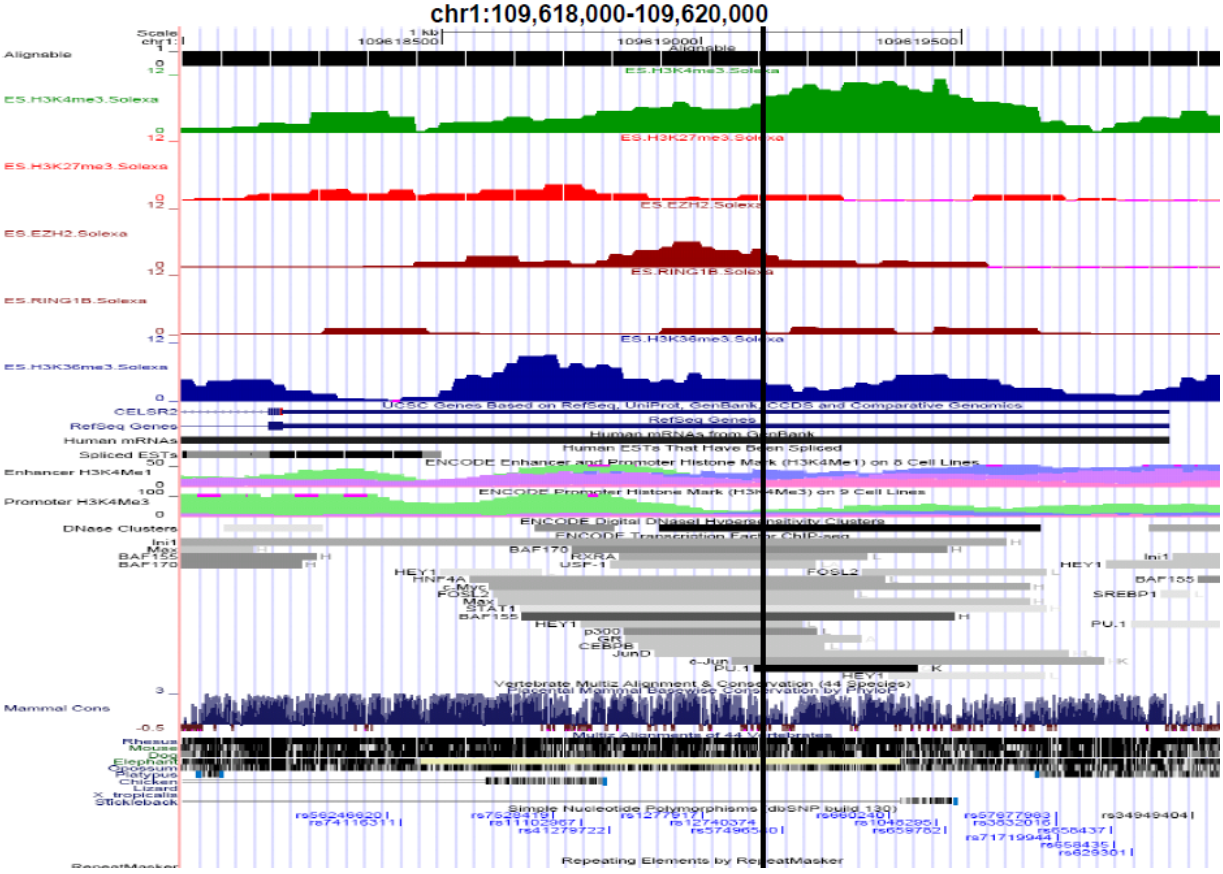
Supplemental Figure S4

Chromatin State Signatures of 43 IDAGL

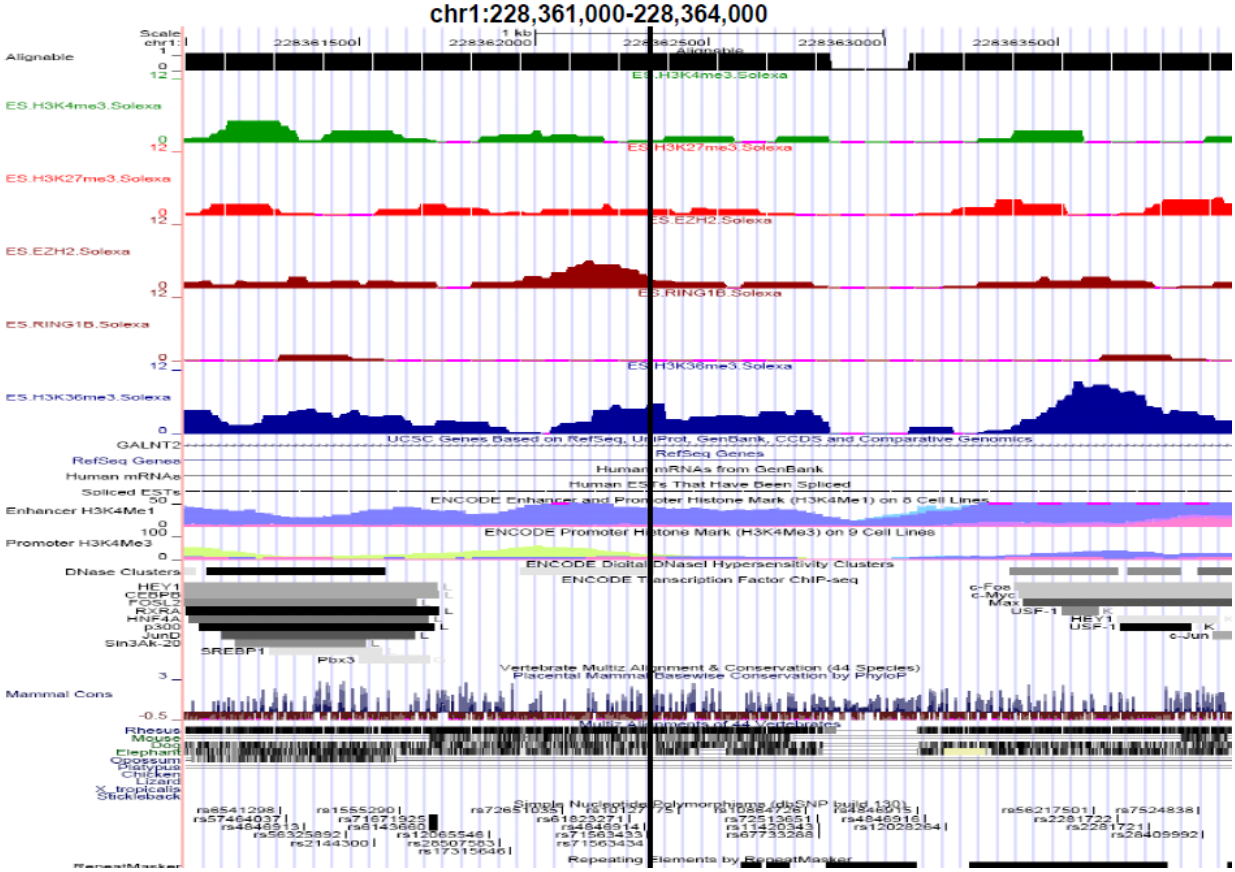

Chromatin state map signatures of IDAGL
Instances
25
20
15
10
5
0
LDL/TC
HDL/TG
HDL/TC
TG
LDL/TC/HDL
LDL/TC
TG/HDL
CAD
CAD
T2D
T2D
T2D
T1D
T1D
RA/T1D
RA
RA
UC
CD
CD
CD
CD
CD
PC, CRC
PC
NAFLP1
BC
PD
PD
PD
SCZH
LONGEVITY
LONGEVITY
LONGEVITY
T2D
CD
CD
BC
BC
MEL
RA
HT
HDL
GR
Ini1
Brg1
BAF170
MAX
c-MYC
RXRA
HNF4A
EBF
IRF4
BATF
PU.1
CTCF
Rad21
c-FOS
CEBP
TAF1
FOSL2
BAF155
p300
c-JUN
JUND
HEY1
STAT1
NFkB
Polycomb
H3K27me3
H3K36Me3
H3K4Me3
H3K4Me1
Enhancer Function
transRNA


Transcription factors map of IDAGL
Instances
18
16
14
12
10
8
6
4
2
0
LDL/TC
HDL/TG
HDL/TC
TG
LDL/TC/HDL
LDL/TC
TG/HDL
CAD
CAD
T2D
T2D
T2D
T1D
T1D
RA/T1D
RA
RA
UC
CD
CD
CD
CD
CD
PC; CRC
PC
NALP1
BC
PD
PD
PD
SCZH
LONGEVITY
LONGEVITY
LONGEVITY
GR
Ini1
Brg1
BAF170
MAX
c-MYC
RXRA
HNF4A
EBF
IRF4
BATF
PU.1
CTCF
Rad21
c-FOS
CEBP
TAF1
FOSL2
BAF155
p300
c-JUN
JUND
HEY1
STAT1
NFkB
transRNA

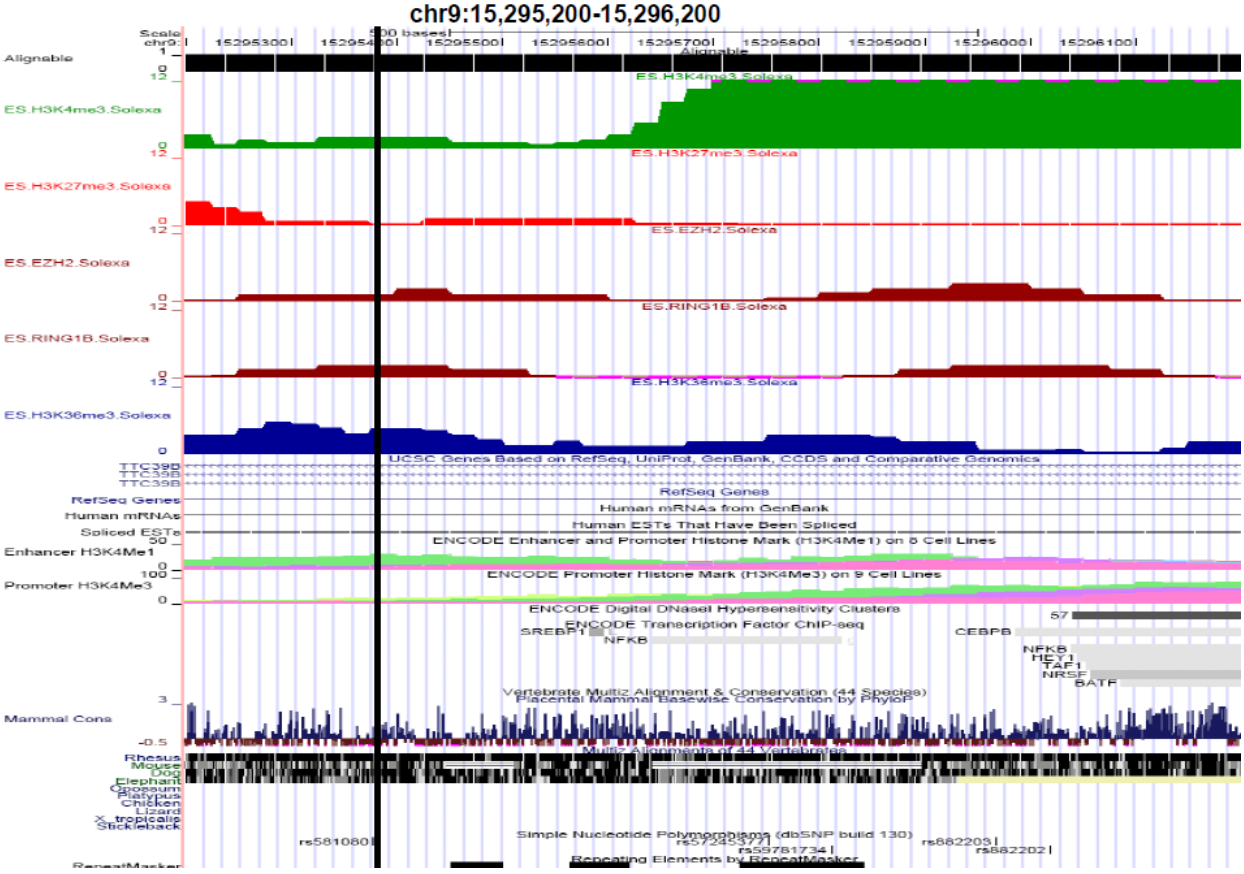
LDL/TC rs12740374
size 2,001 bp.
chr1:109,618,000-109,620,000



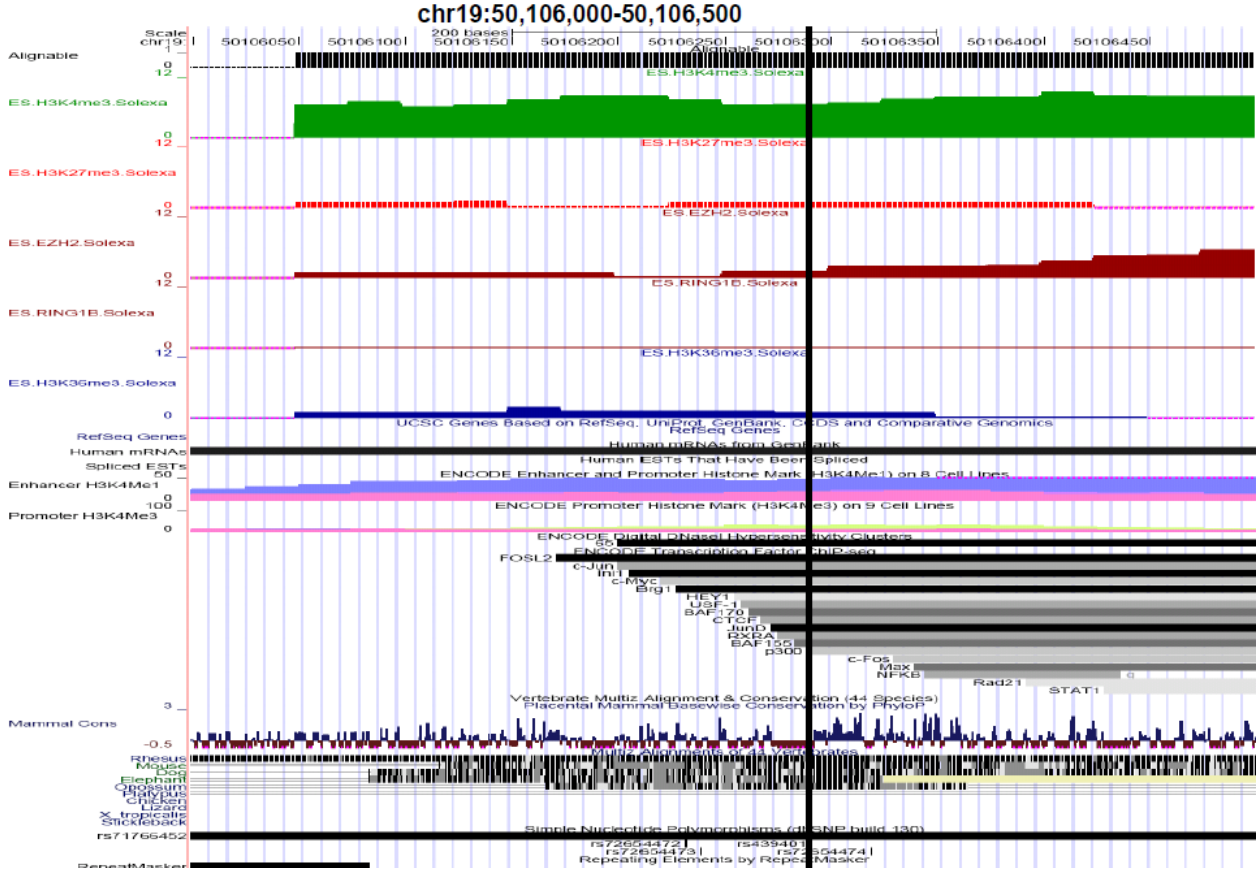
HDL/TG rs4846914 Intronic GALNT2
size 3,001 bp.
chr1:228,361,000-228,364,000



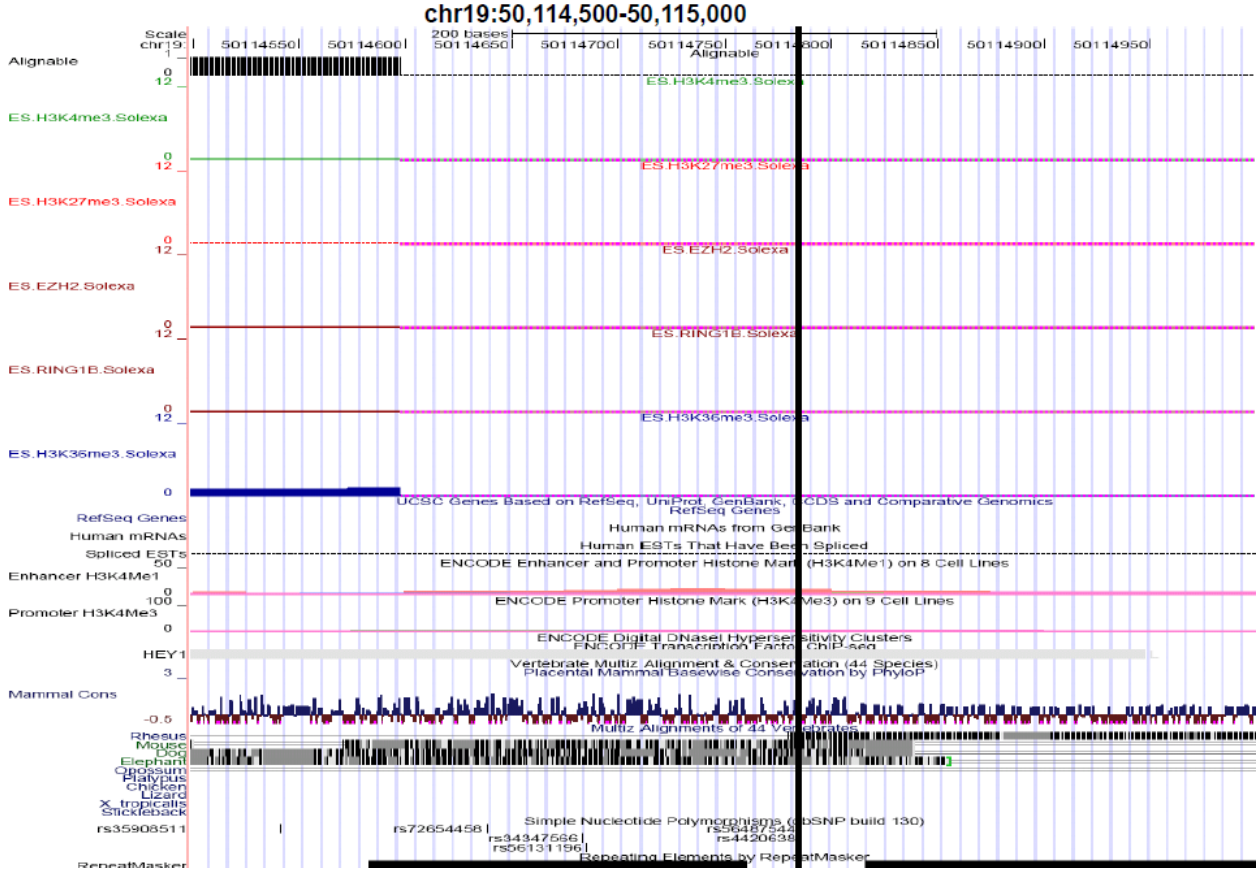
HDL/TC TTC39B rs581080
size 1,001 bp.
chr9:15,295,200-15,296,200



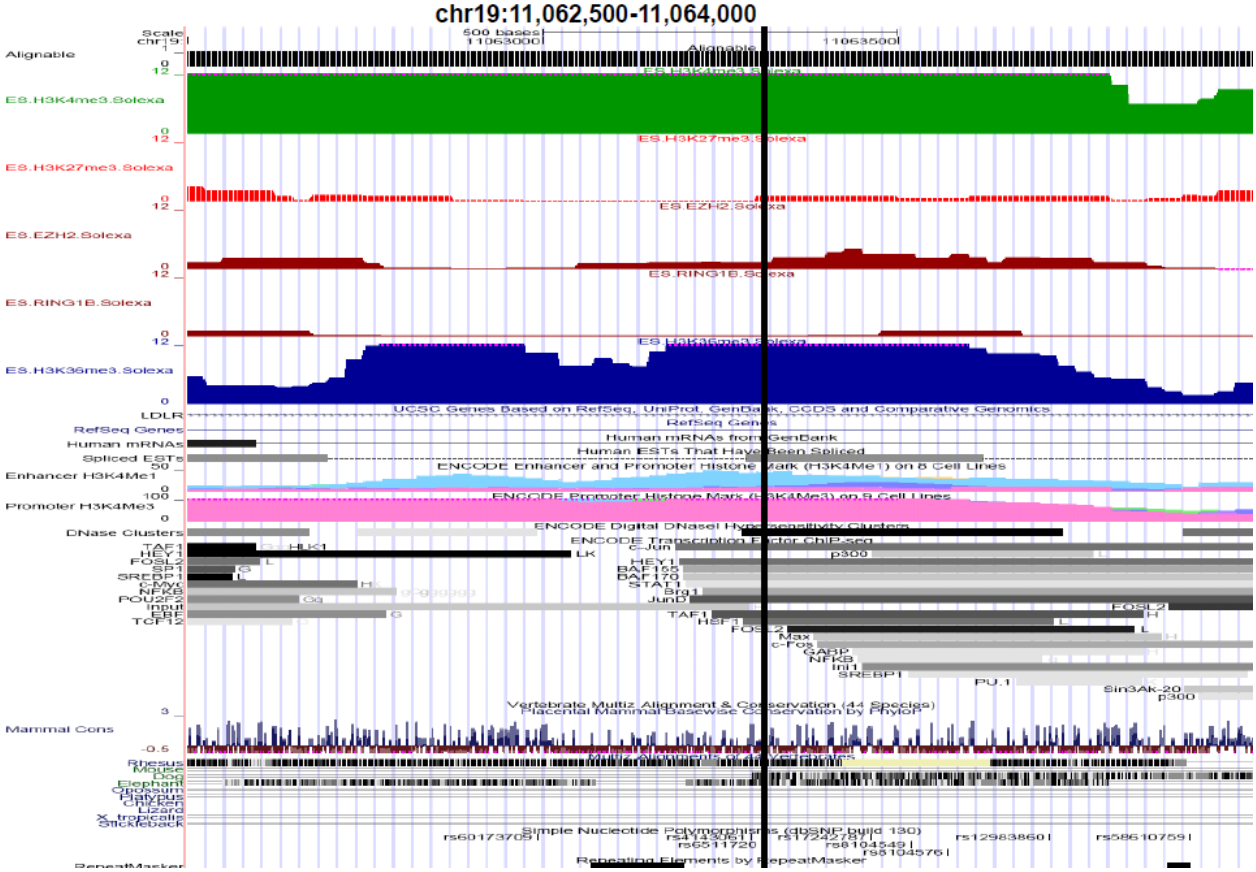
TG rs439401
size 501 bp.
chr19:50,106,000-50,106,500



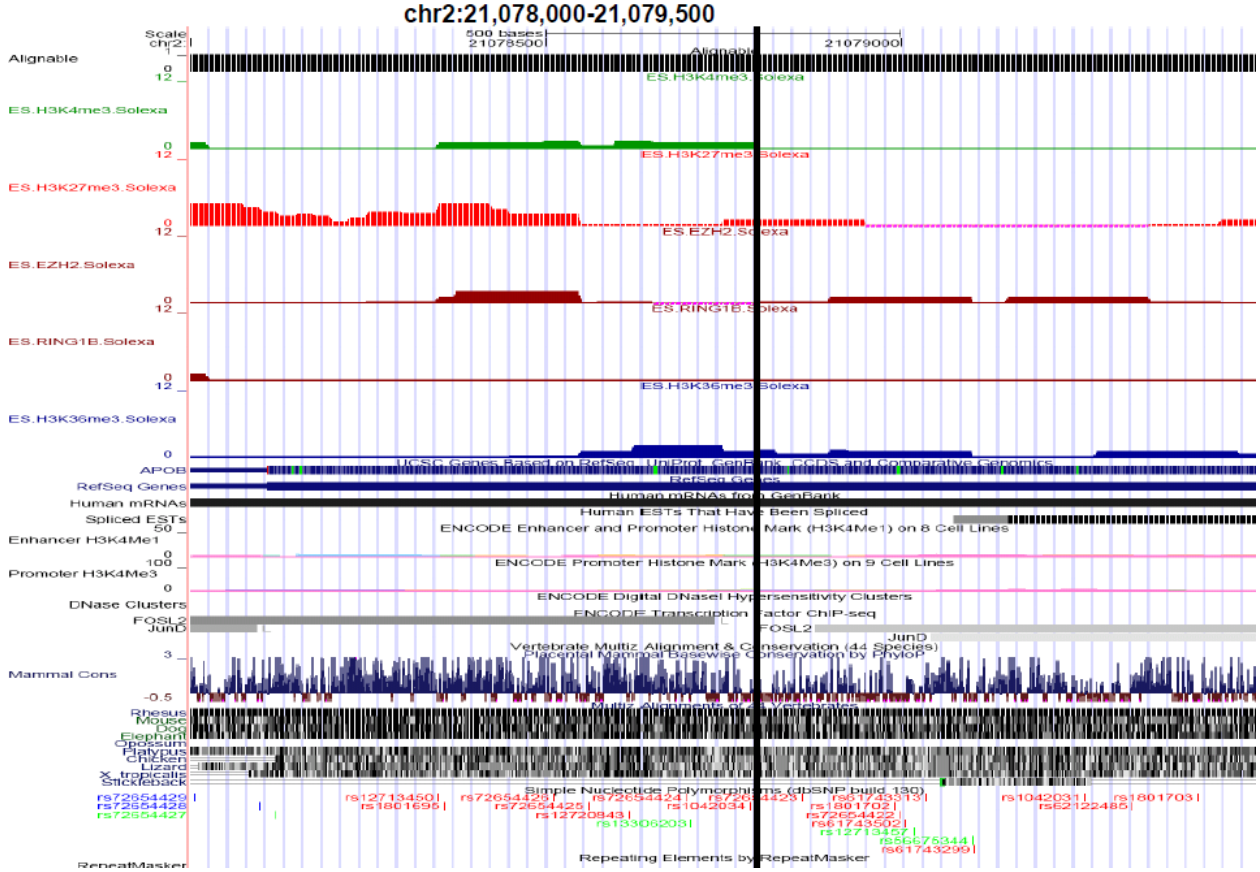
LDL/TC/HDL rs4420638
size 501 bp
chr19:50,114,500-50,115,000



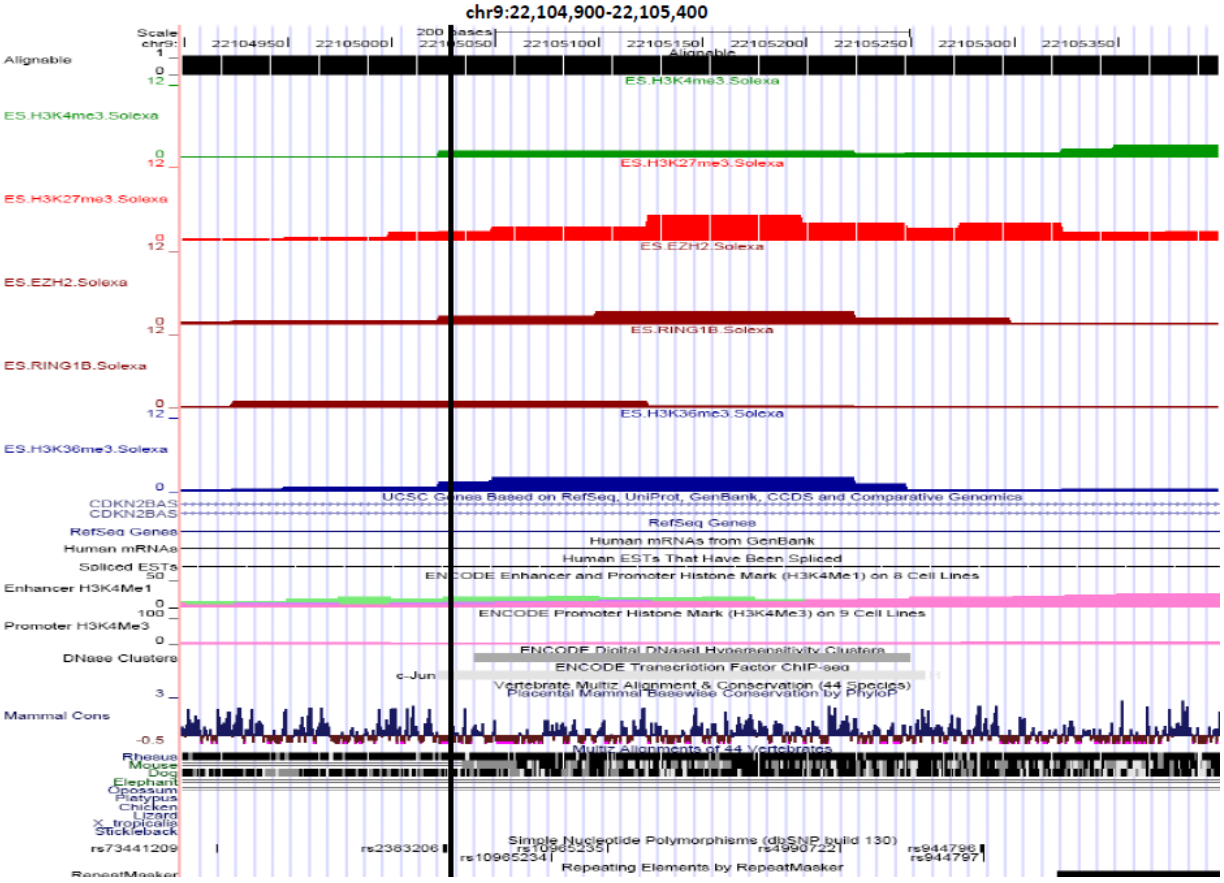
LDL/TC rs651720 LDLR
size 1,501 bp.
chr19:11,062,500-11,064,000



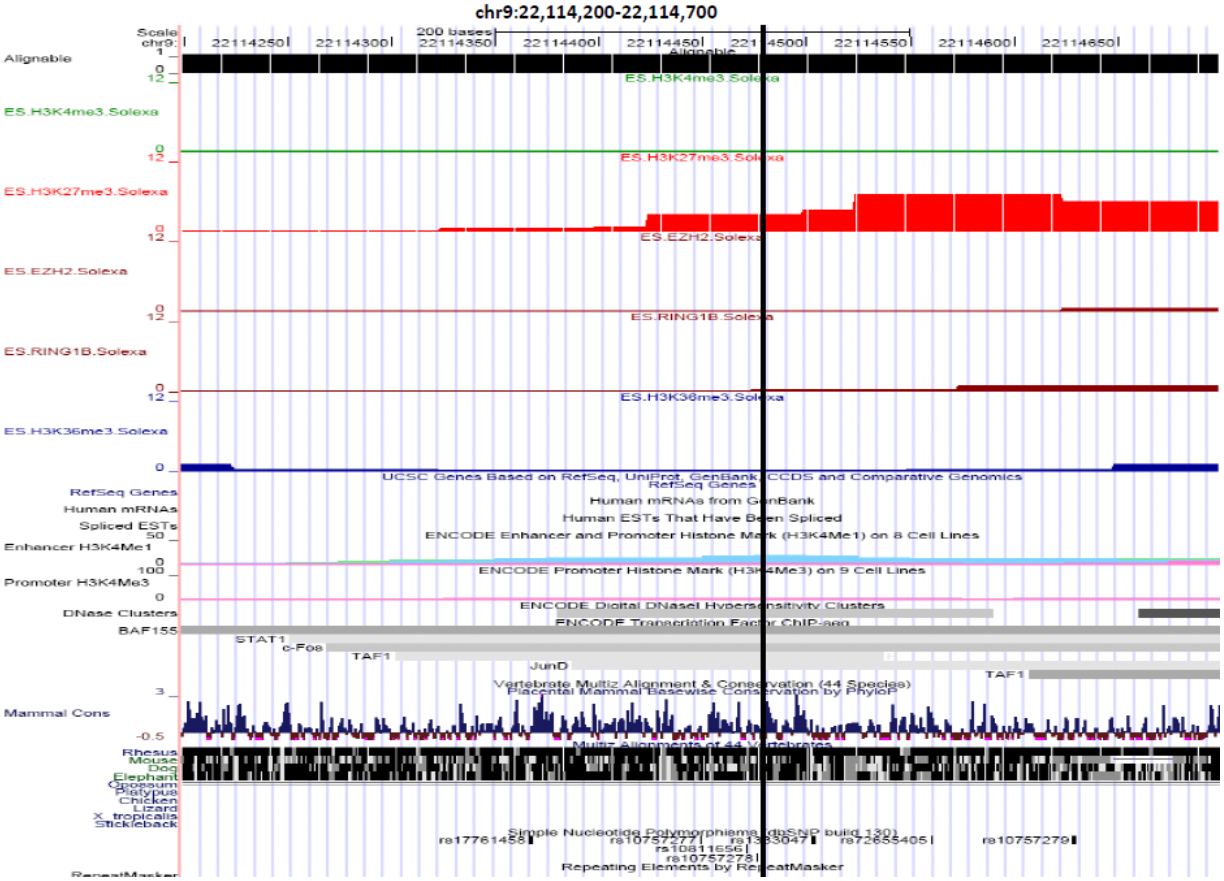
TG/HDL rs1042034
size 1,501 bp.
chr2:21,078,000-21,079,500



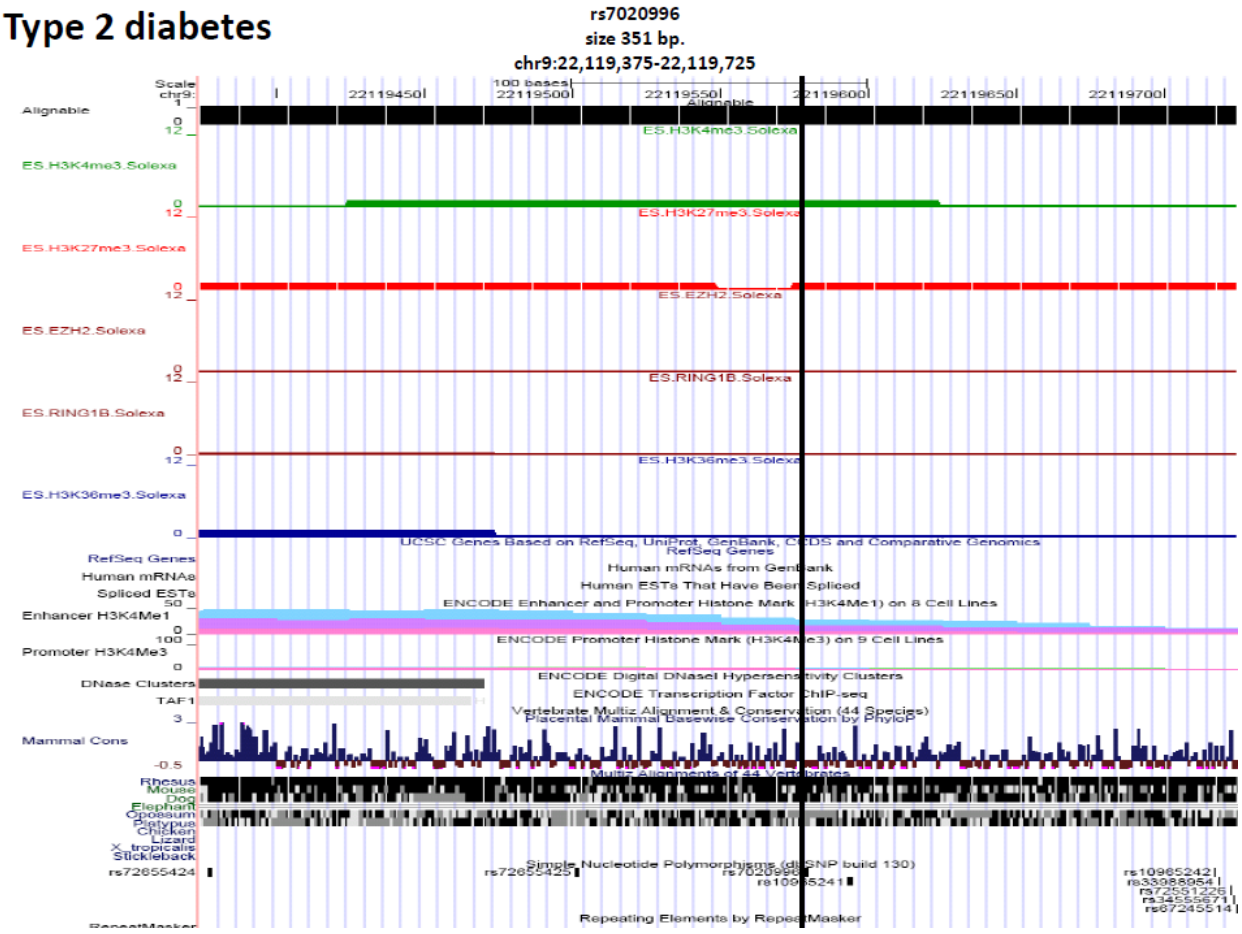
rs2383206 CAD
size 502 bp.
chr9:22,104,900-22,105,400



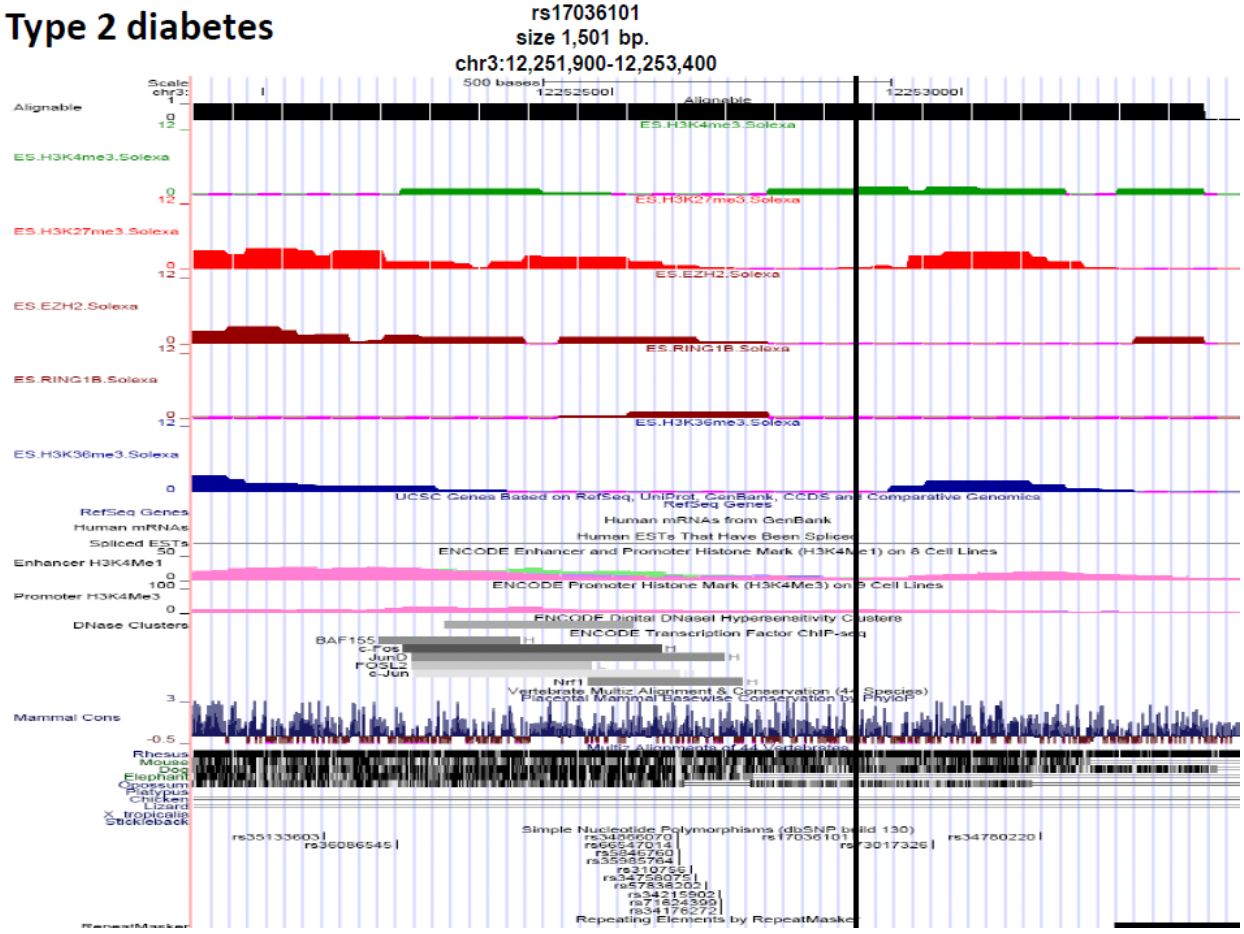
rs10757278 CAD
size 501 bp.
chr9:22,114,200-22,114,700



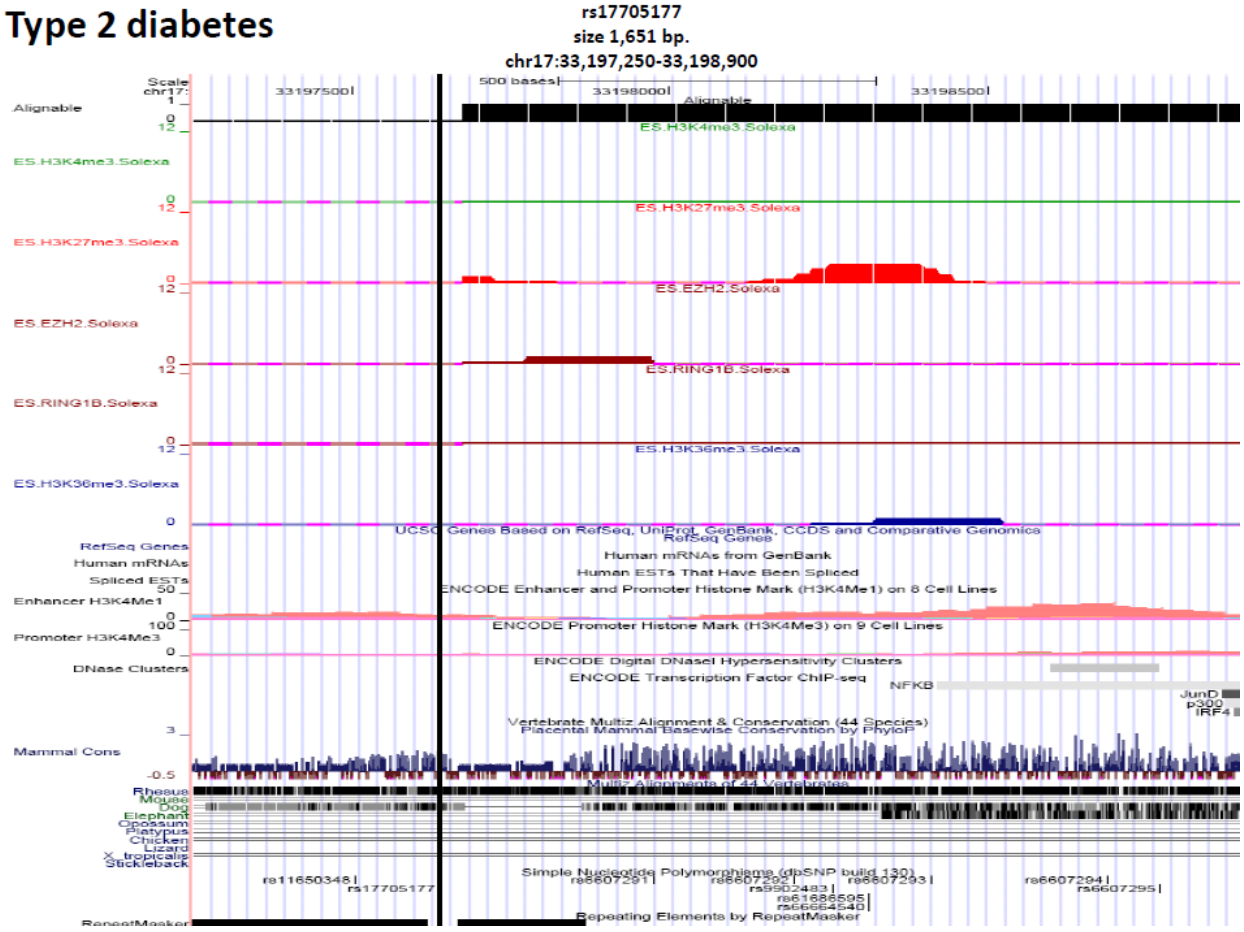
Type 2 diabetes



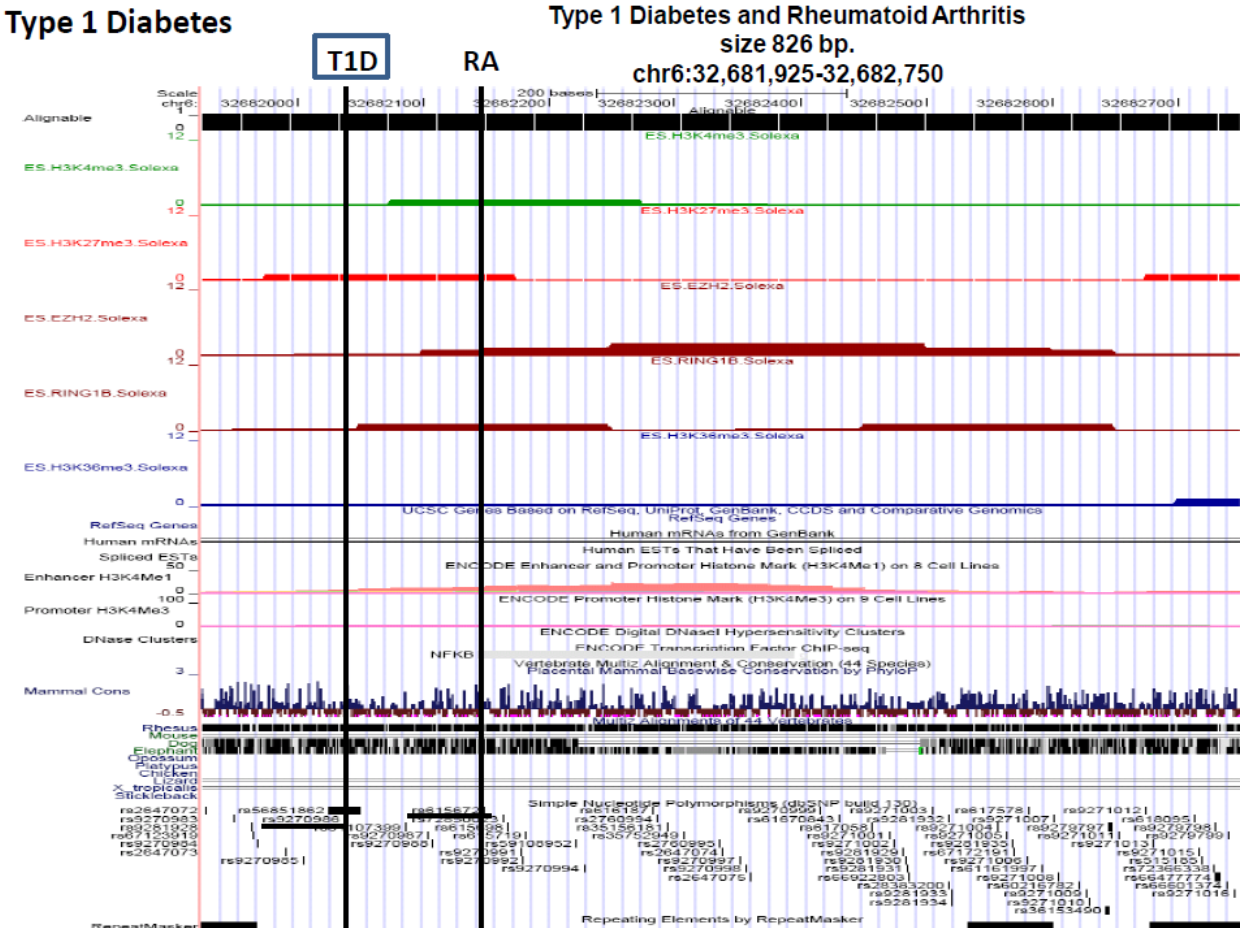
Type 2 diabetes



Type 2 diabetes

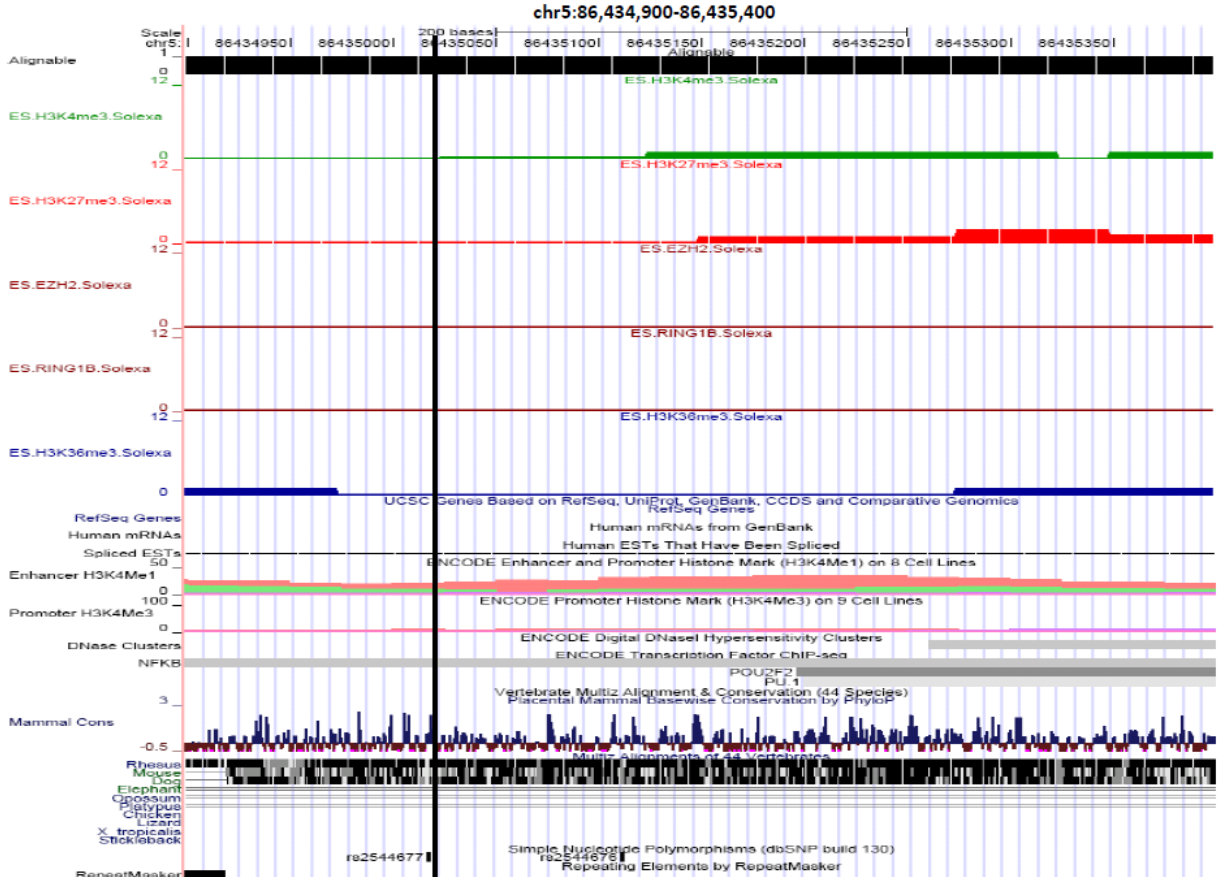


Type 1 Diabetes

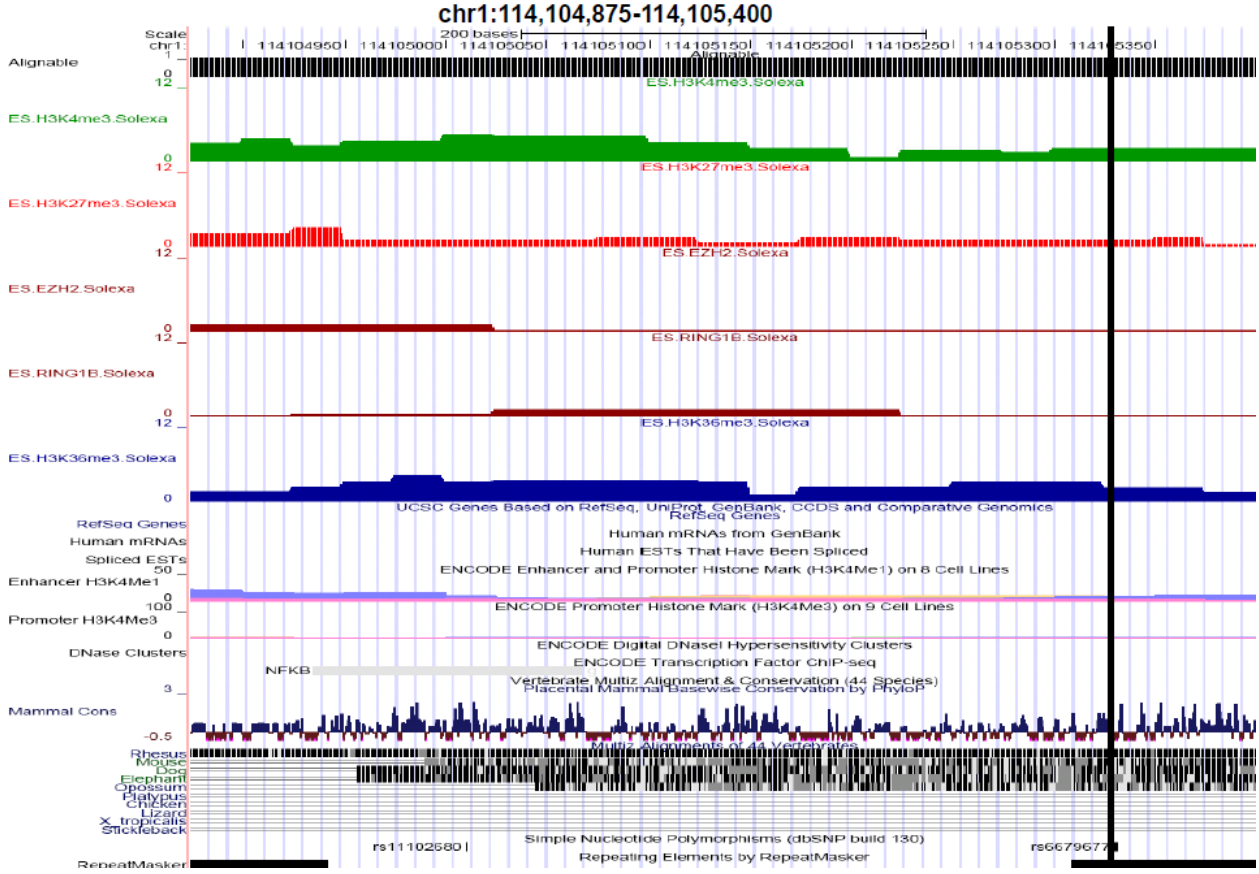


Type 1 Diabetes

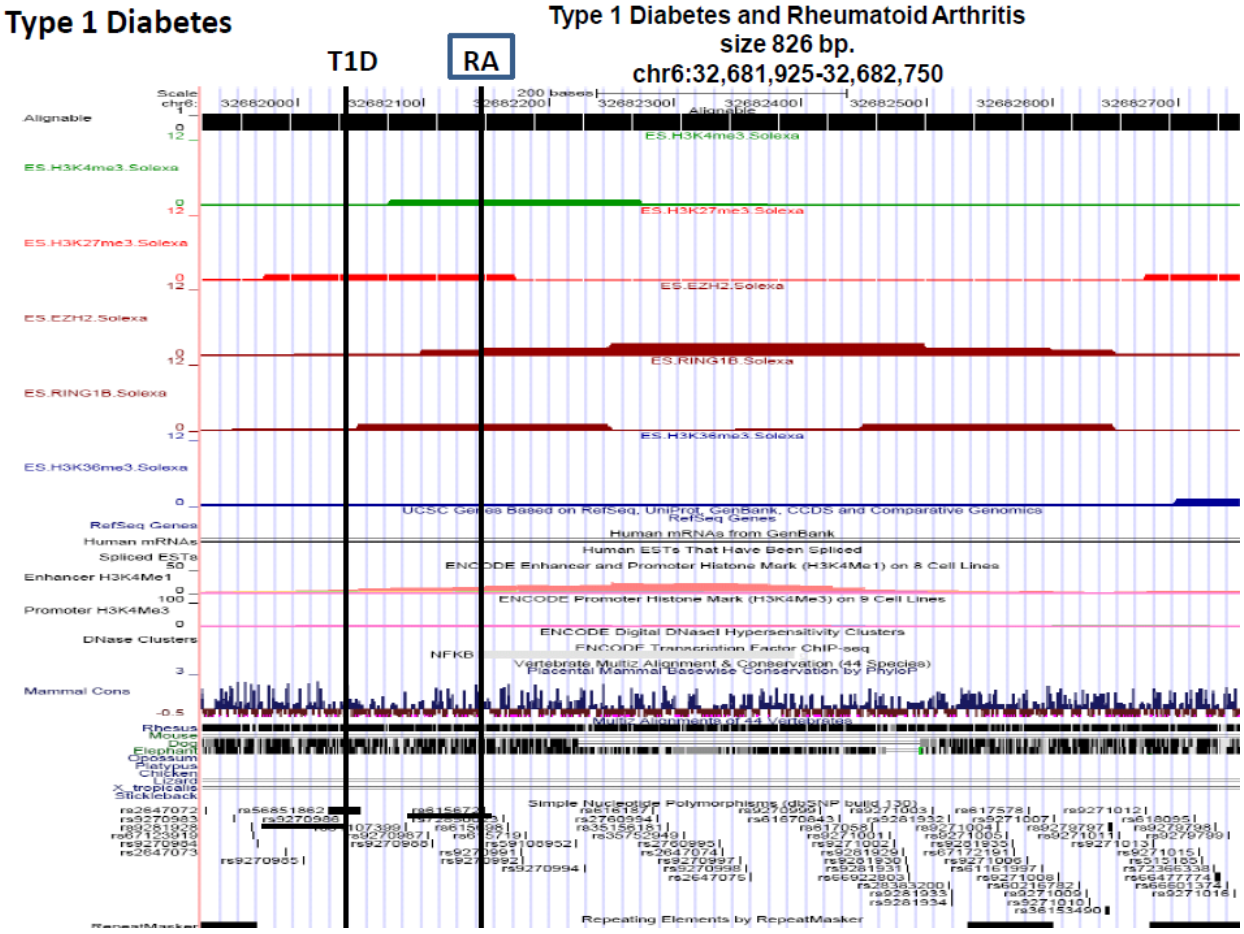
rs2544677
size 501 bp.
chr5:86,434,900-86,435,400



RA and T1D rs6679677
size 526 bp.
chr1:114,104,875-114,105,400

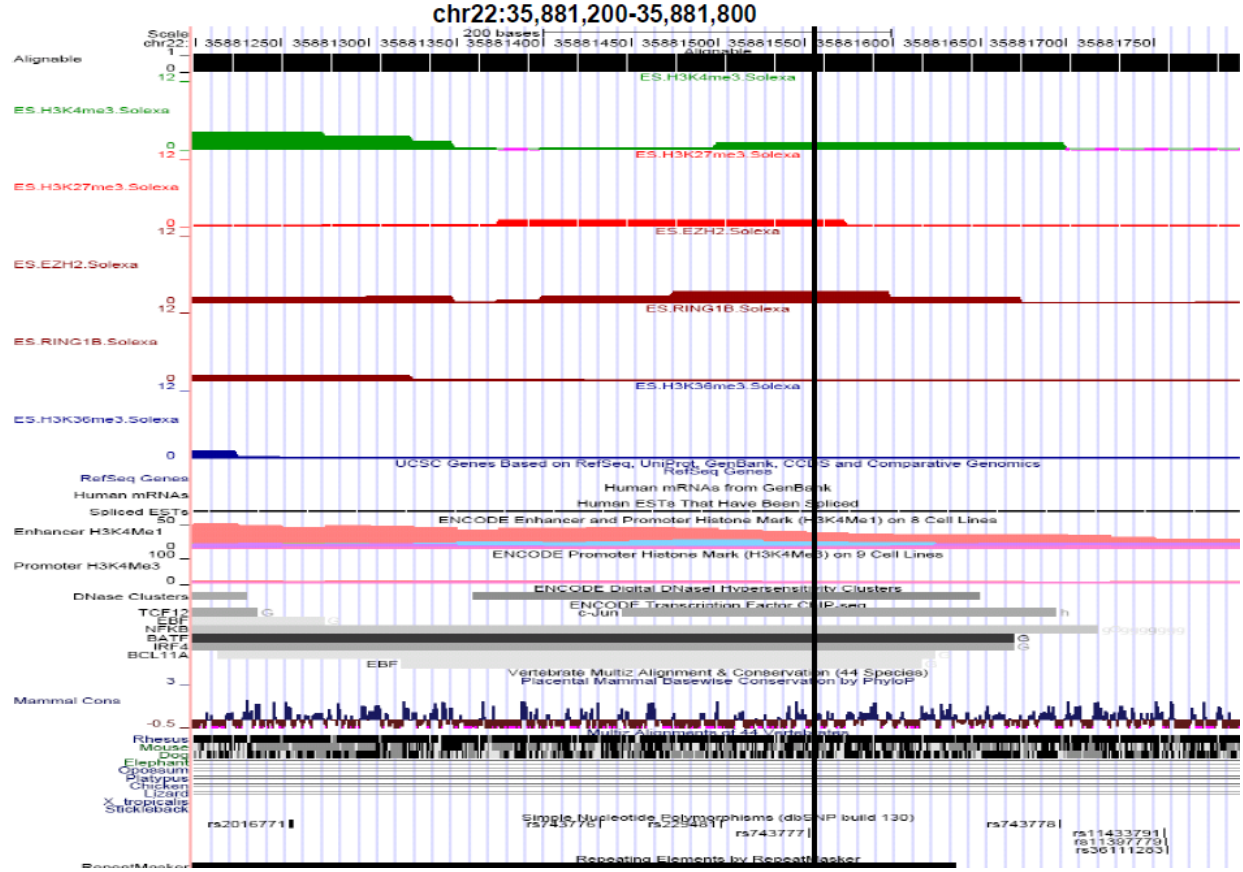


Type 1 Diabetes



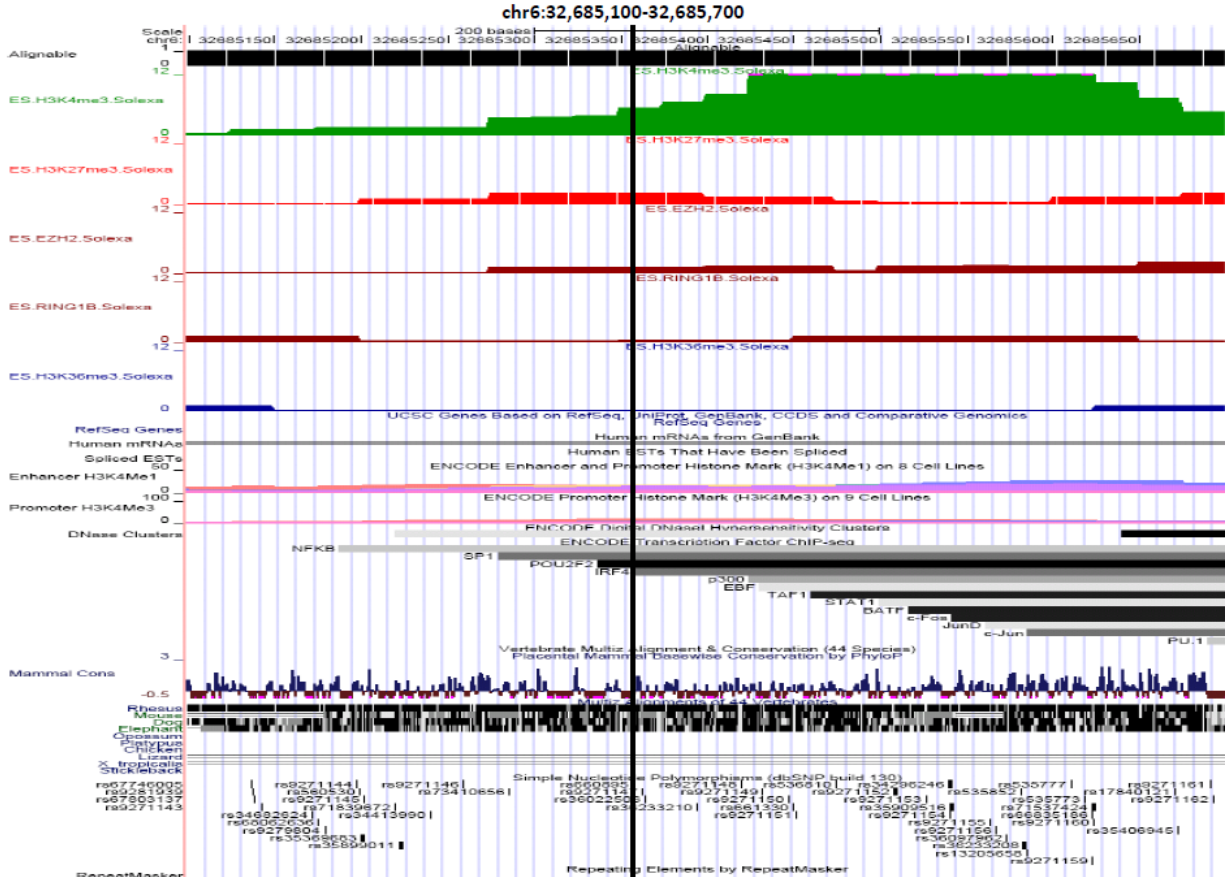
Rheumatoid Arthritis

rs743777 RA Enhancer
size 601 bp.
chr22:35,881,200-35,881,800

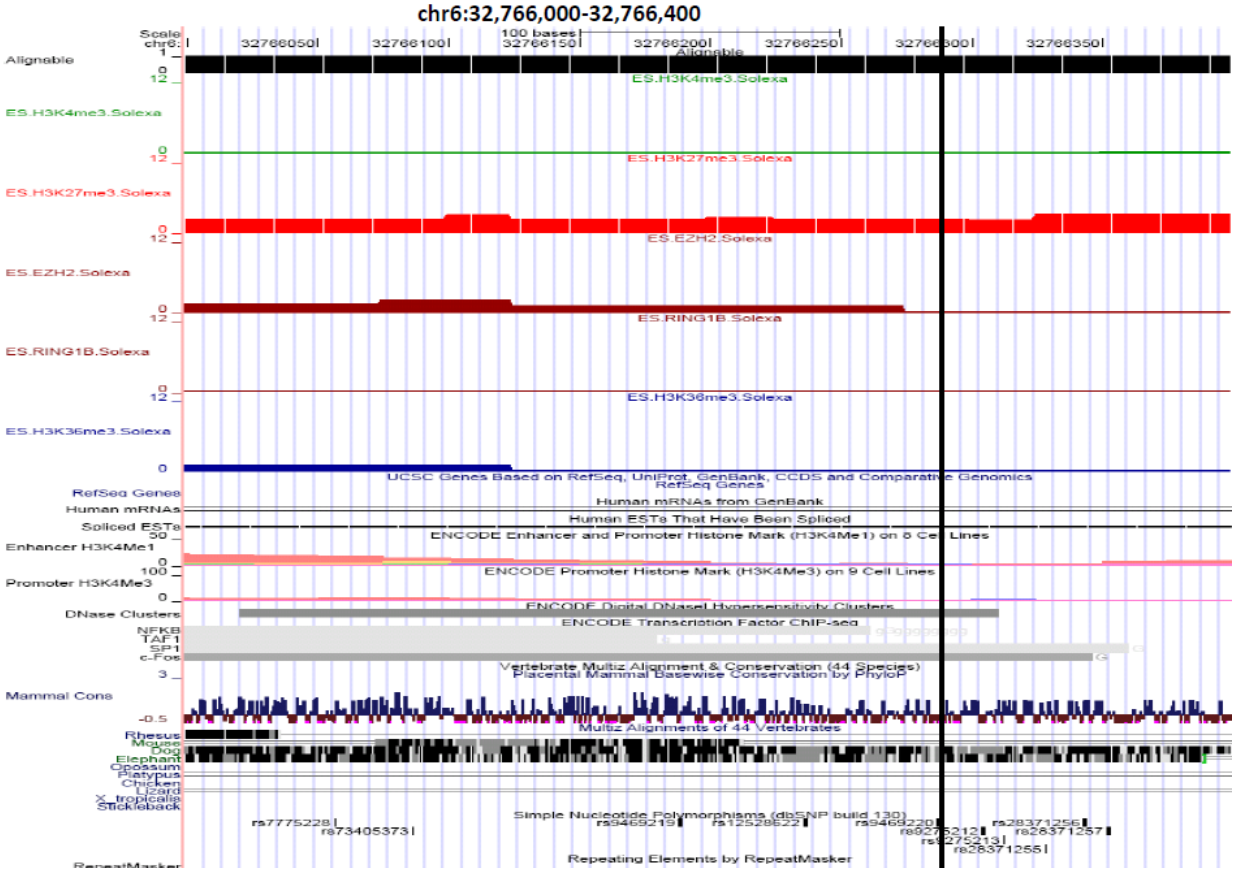


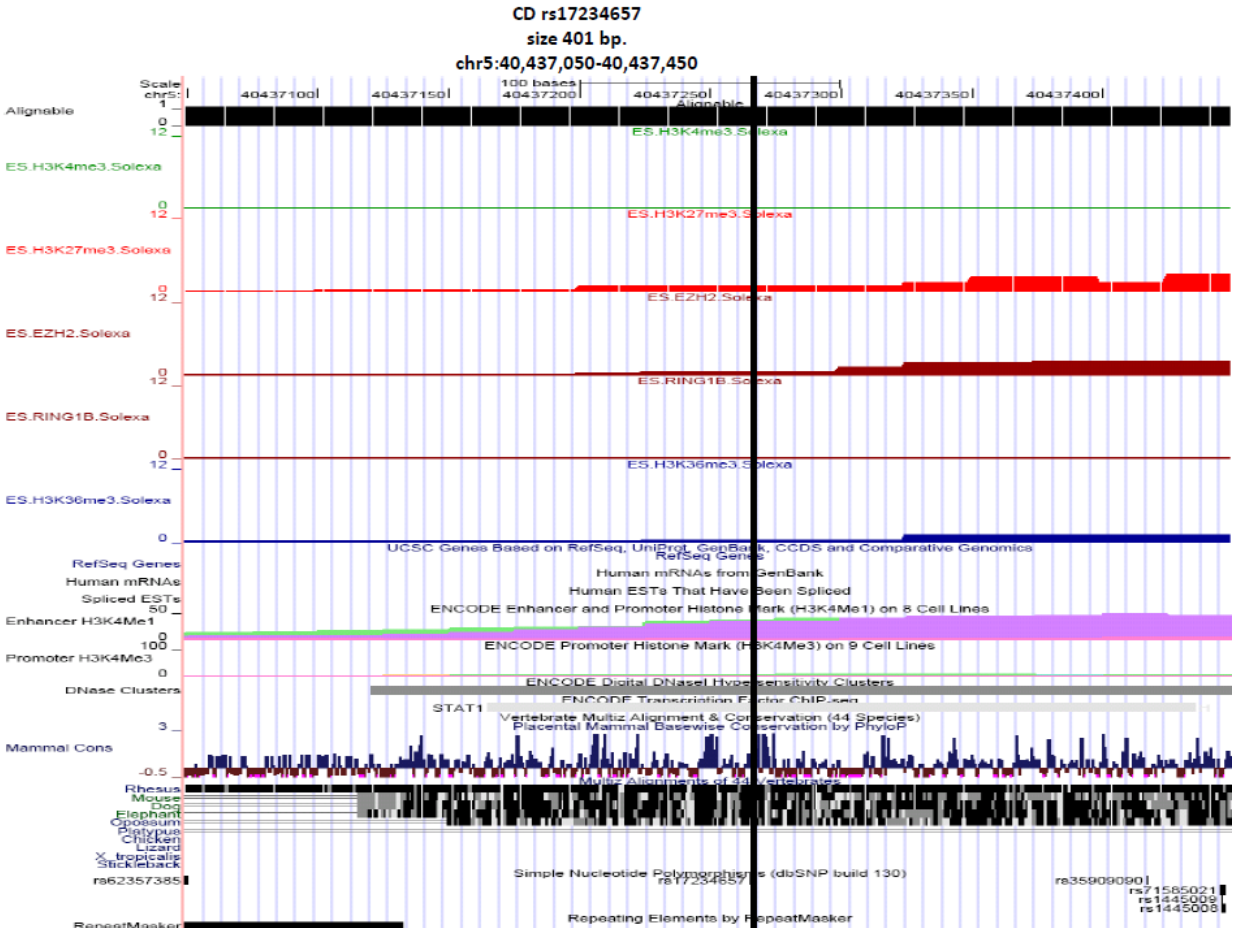
Ulcerative colitis

rs660895
size 601 bp.
chr6:32,685,100-32,685,700

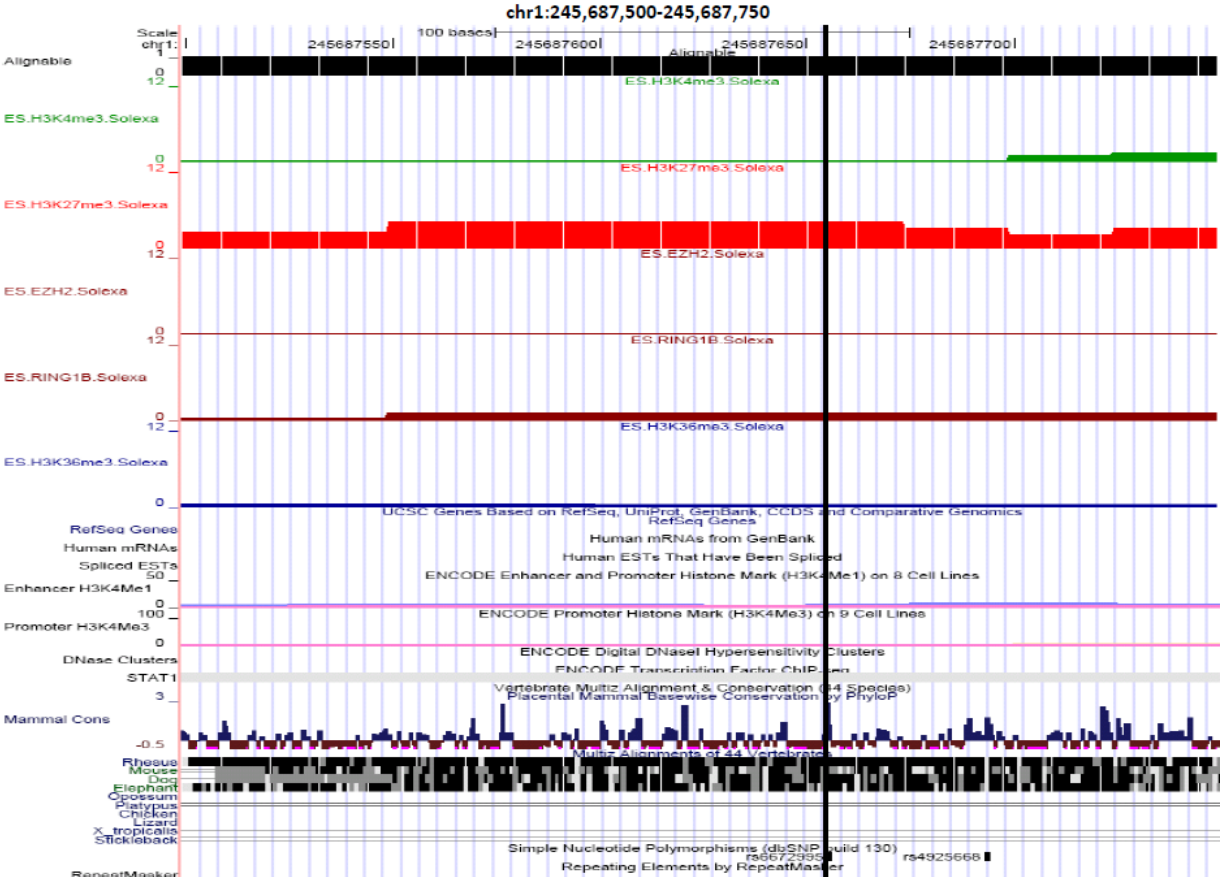


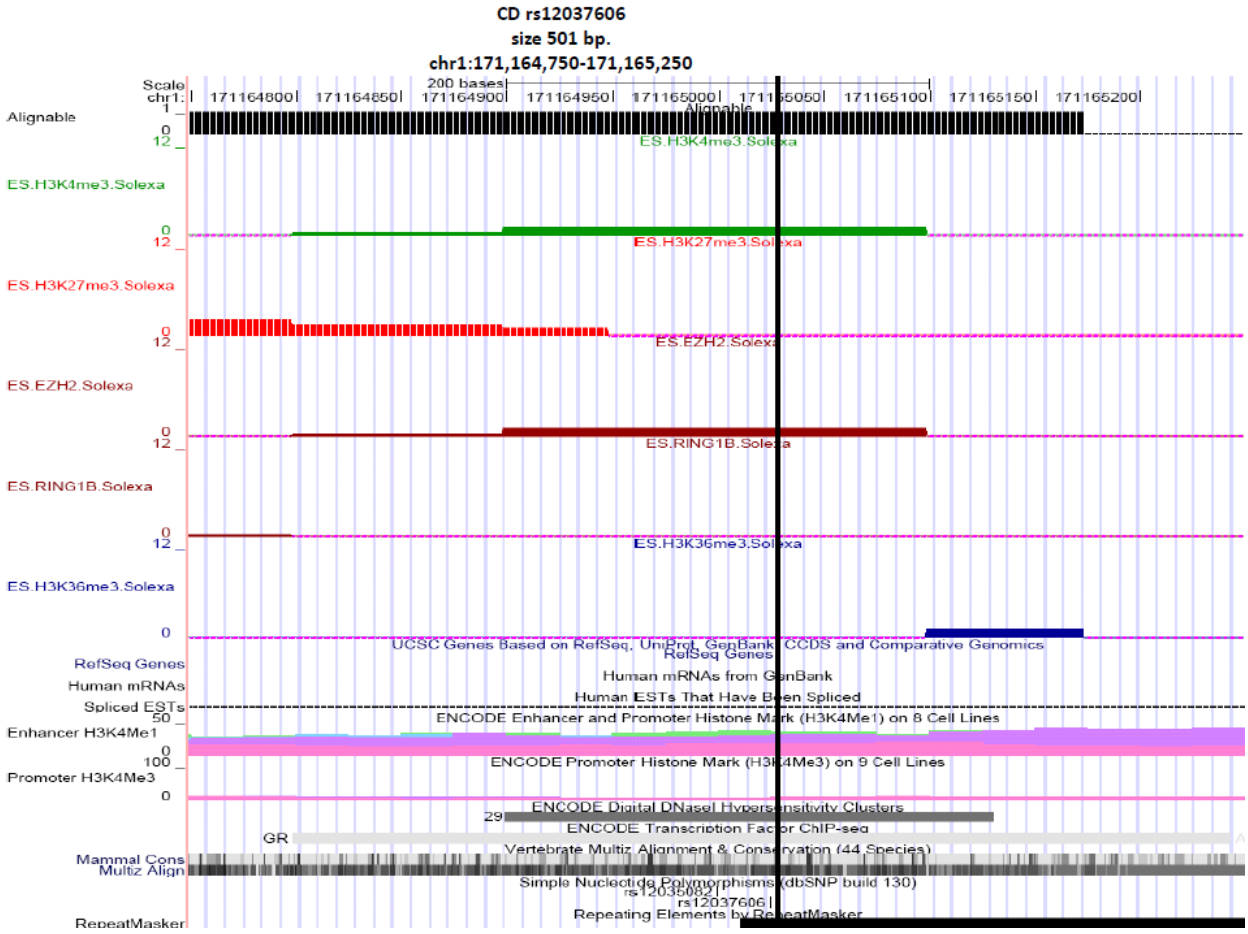
CD rs9469220
size 401 bp.
chr6:32,766,000-32,766,400



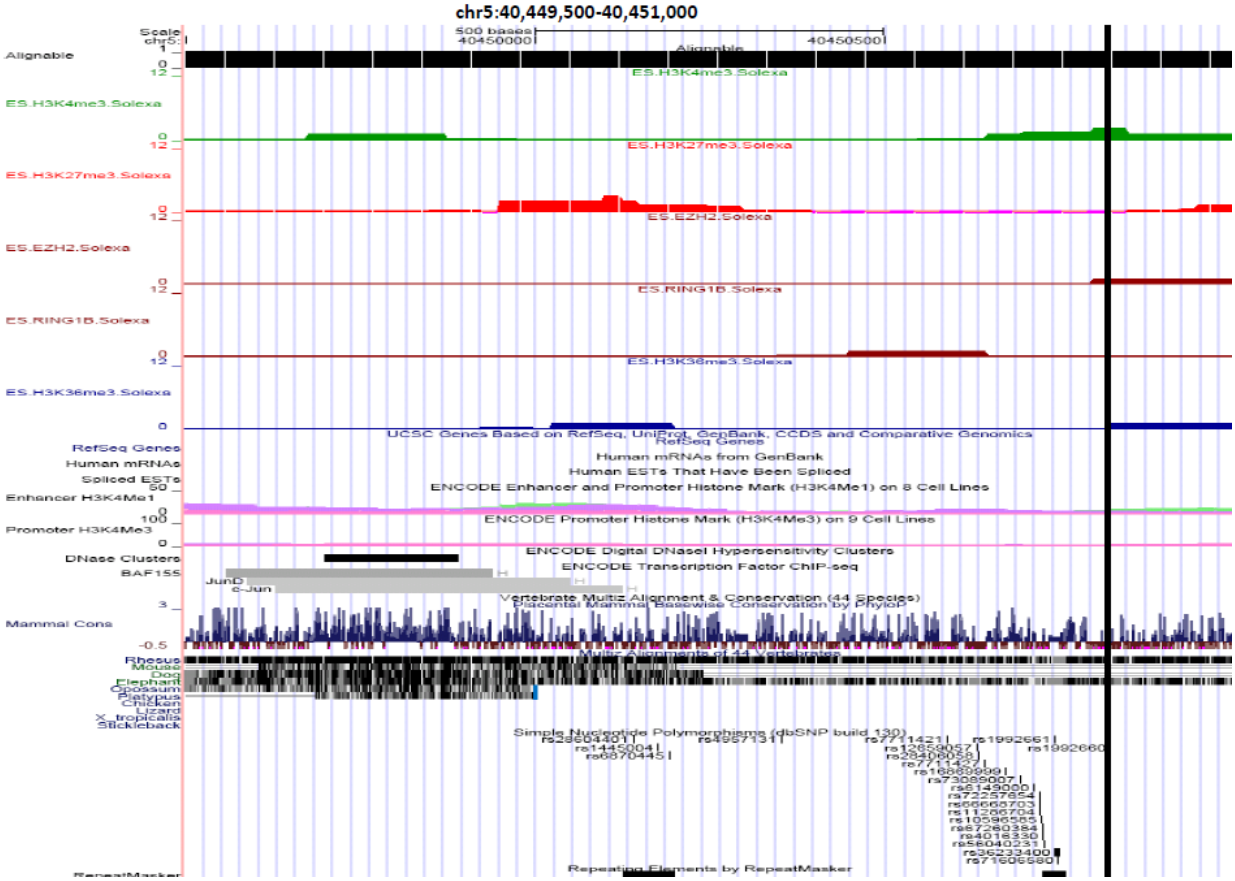

CD rs17234657
size 401 bp.
chr5:40,437,050-40,437,450
Scale
chr5:
100 bases
40437100
40437150
40437200
40437250
40437300
40437350
40437400
Alignable
ES.H3K4me3.Solexa
ES.H3K27me3.Solexa
ES.EZH2.Solexa
ES.RING1B.Solexa
ES.H3K36me3.Solexa
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
RefSeq Genes
Human mRNAs
Human mRNAs from GenBank
Spliced ESTs
Human ESTs That Have Been Spliced
Enhancer H3K4Me1
ENCODE Enhancer and Promoter Histone Mark (H3K4Me1) on 8 Cell Lines
Promoter H3K4Me3
ENCODE Promoter Histone Mark (H3K4Me3) on 9 Cell Lines
DNase Clusters
ENCODE Digital DNaseI Hypersensitivity Clusters
STAT1
Mammal Cons
Vertebrate Multiz Alignment & Conservation (44 Species)
Placental Mammal Basewise Conservation by PhyloP
Rhesus
Mouse
Dog
Elephant
Opossum
Platypus
Chicken
Lizard
X. tropicalis
Stickleback
rs62357385
Simple Nucleotide Polymorphisms (dbSNP build 130)
rs17234657
rs35909090
rs71585021
rs1445009
rs1445008
RepeatMasker
Repeating Elements by RepeatMasker

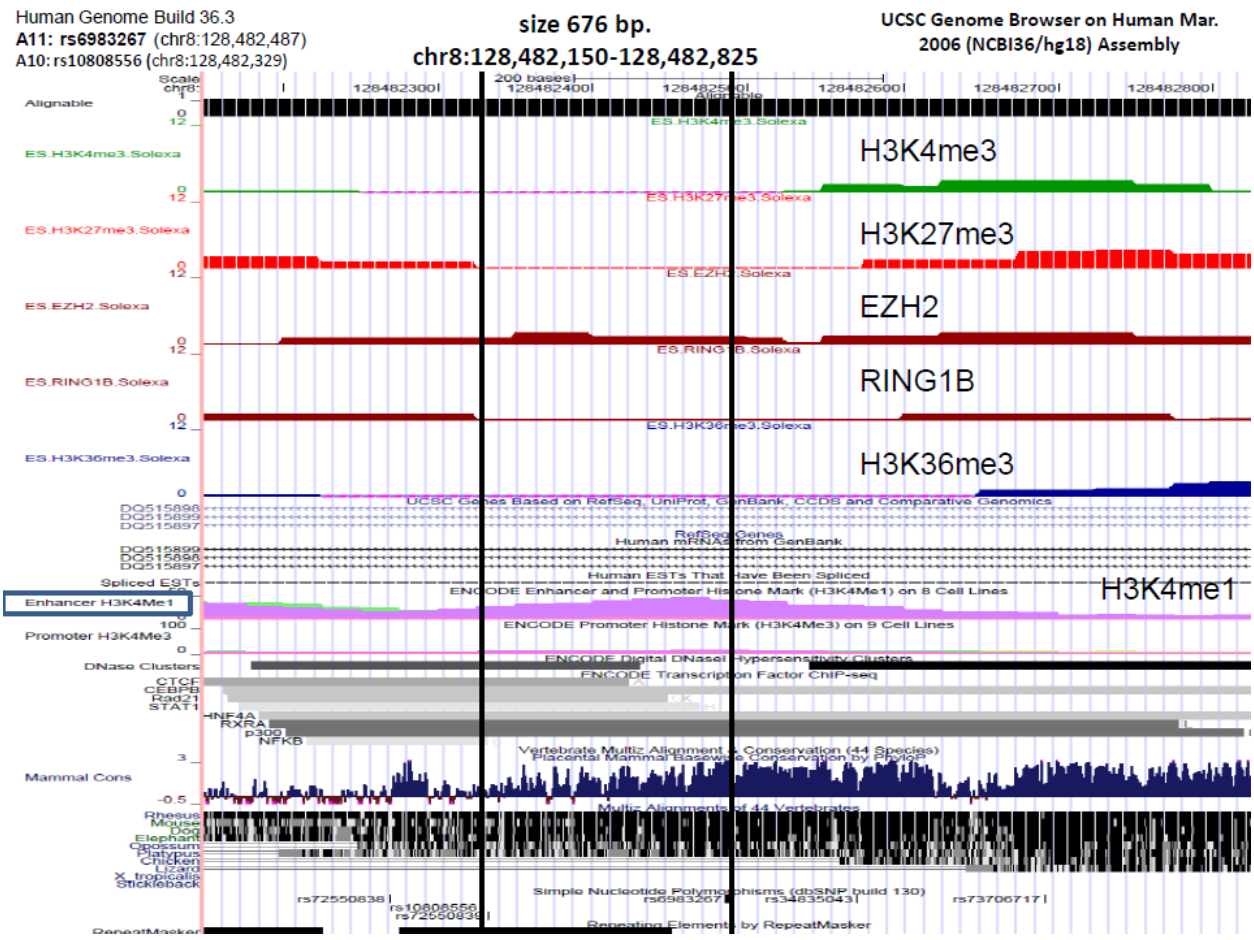
CD rs6672995
size 251 bp.
chr1:245,687,500-245,687,750



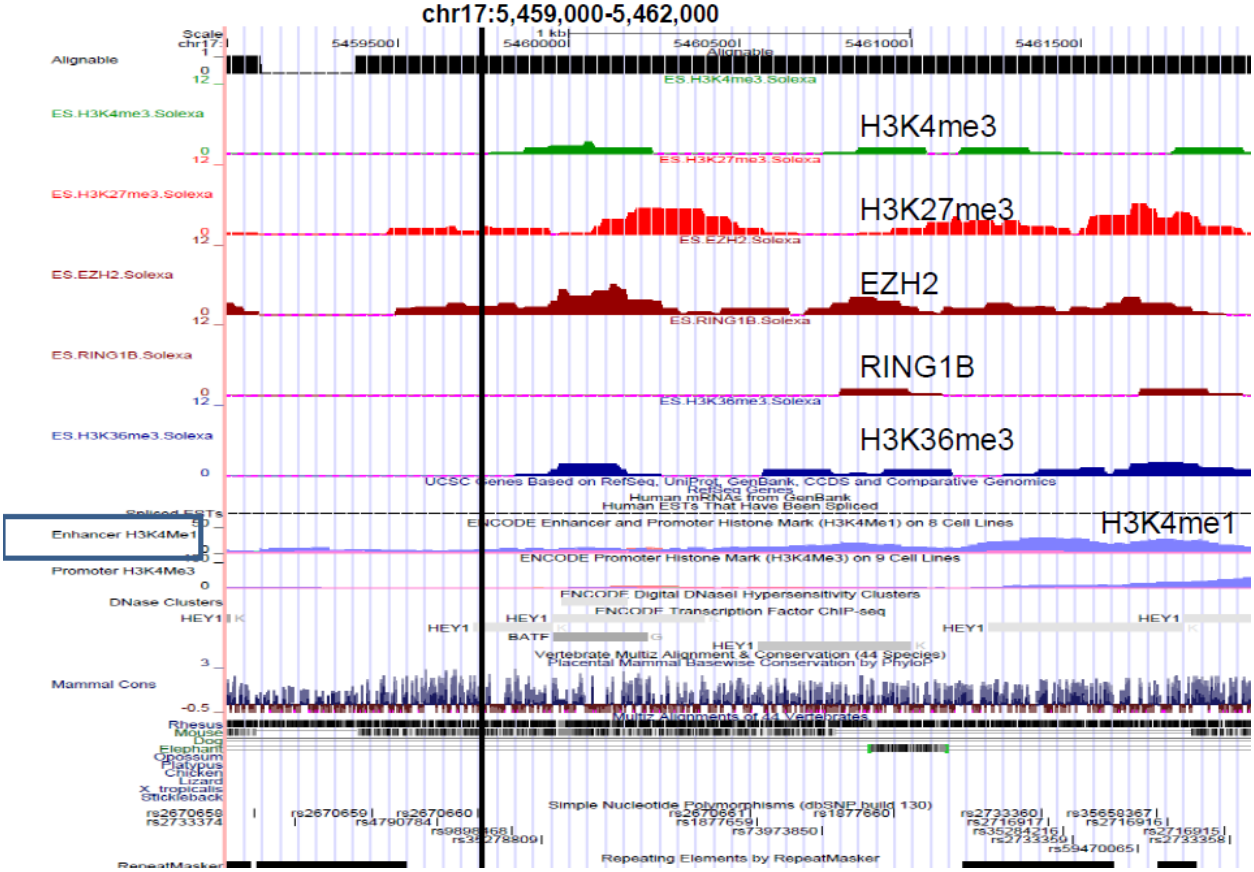

CD rs12037606
size 501 bp.
chr1:171,164,750-171,165,250
Scale
chr1:
200 bases
171164800
171164850
171164900
171164950
171165000
171165050
171165100
171165150
171165200
Alignable
ES.H3K4me3.Solexa
ES.H3K27me3.Solexa
ES.EZH2.Solexa
ES.RING1B.Solexa
ES.H3K36me3.Solexa
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
RefSeq Genes
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
ENCODE Enhancer and Promoter Histone Mark (H3K4Me1) on 8 Cell Lines
ENCODE Promoter Histone Mark (H3K4Me3) on 9 Cell Lines
ENCODE Digital DNaseI Hypersensitivity Clusters
ENCODE Transcription Factor ChIP-seq
GR
Vertebrate Multiz Alignment & Conservation (44 Species)
Simple Nucleotide Polymorphisms (dbSNP build 130)
rs12036082
rs12037606
Repeating Elements by RepeatMasker

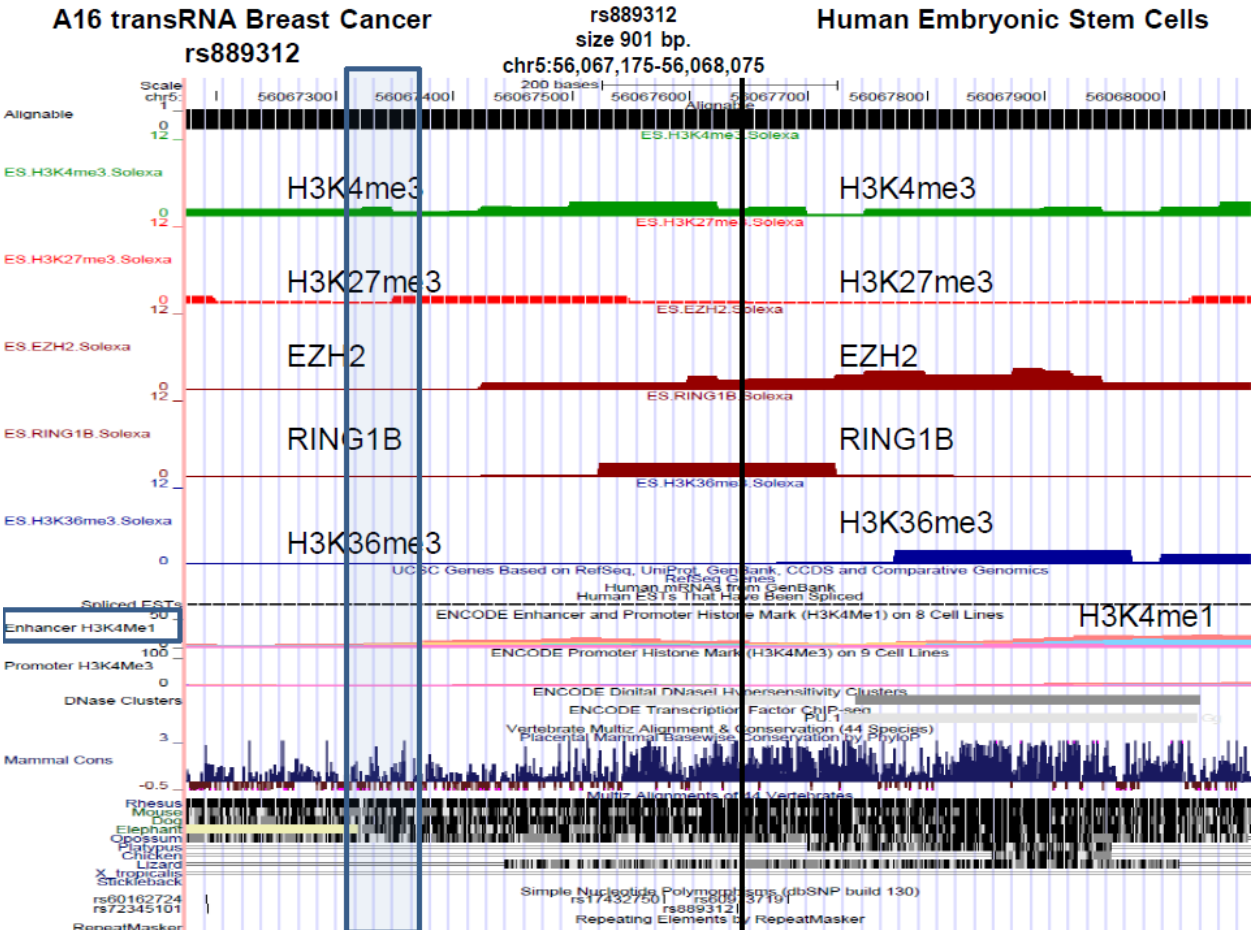
CD rs1992660
size 1,501 bp.
chr5:40,449,500-40,451,000



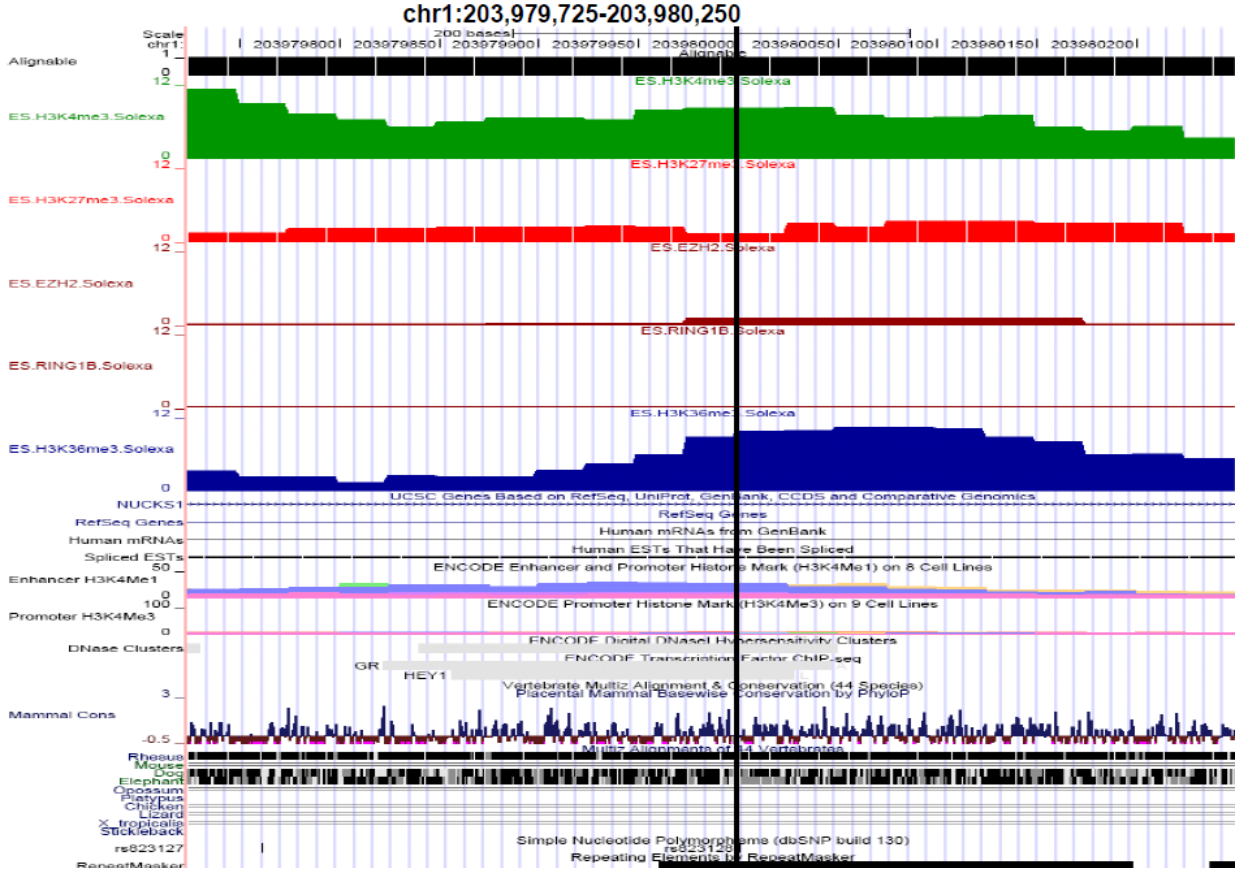

Human Genome Build 36.3
A11: rs6983267 (chr8:128,482,487)
A10: rs10808556 (chr8:128,482,329)
size 676 bp.
chr8:128,482,150-128,482,825
UCSC Genome Browser on Human Mar. 2006 (NCBI36/hg18) Assembly
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
H3K4me1
Enhancer H3K4Me1
Promoter H3K4Me3
Mammal Cons
Simple Nucleotide Polymorphisms (dbSNP build 130)
rs6983267
rs10808556
rs72550838
rs72550839
rs34835043
rs73706717

NALP1-locus transRNA
size 3,001 bp.
chr17:5,459,000-5,462,000

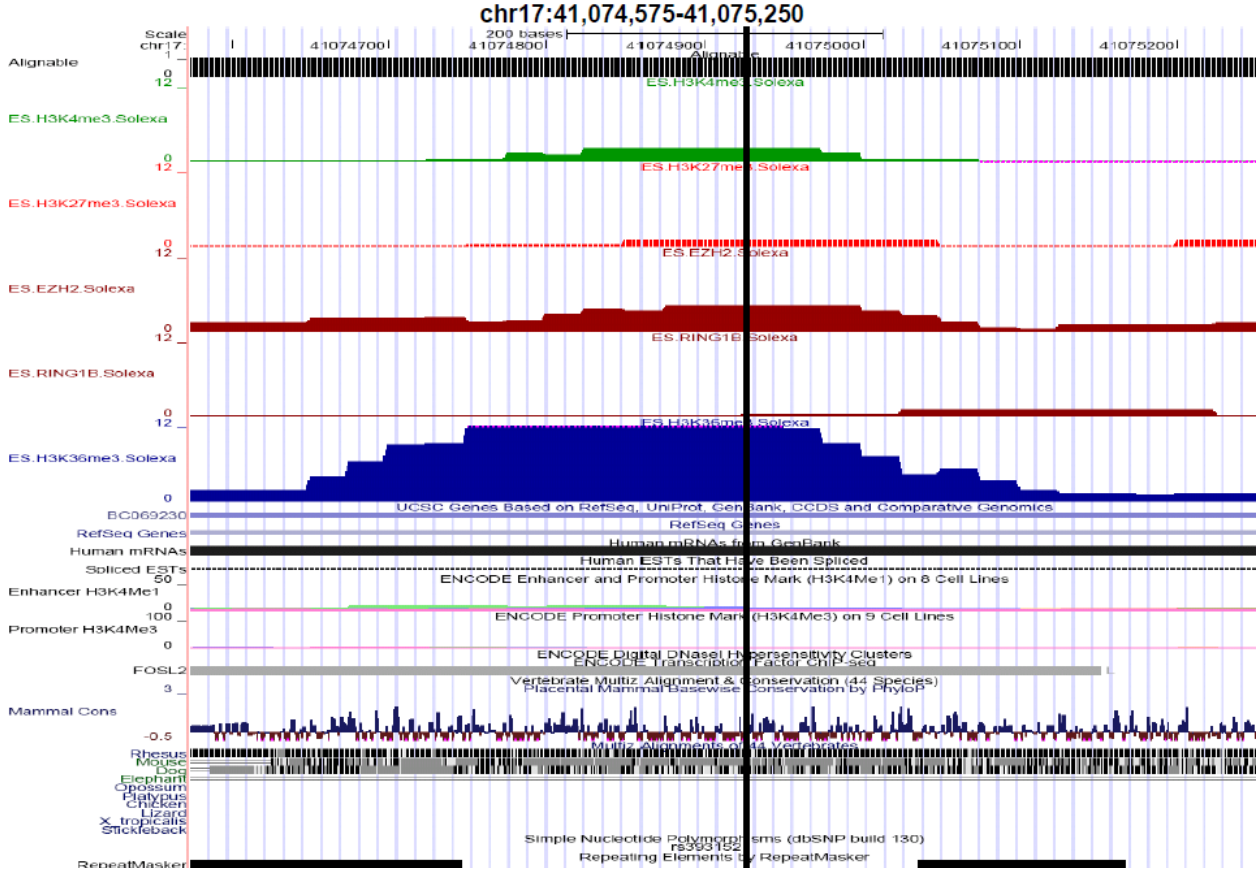



A16 transRNA Breast Cancer
rs889312
rs889312
size 901 bp.
chr5:56,067,175-56,068,075
Human Embryonic Stem Cells
H3K4me3
H3K27me3
EZH2
RING1B
H3K36me3
H3K4me1
Alignable
ES.H3K4me3.Solexa
ES.H3K27me3.Solexa
ES.EZH2.Solexa
ES.RING1B.Solexa
ES.H3K36me3.Solexa
Enhancer H3K4Me1
Promoter H3K4Me3
DNase Clusters
Mammal Cons
Rhesus
Mouse
Dog
Elephant
Opossum
Platypus
Chicken
Lizard
X. tropicalis
Stickleback
rs60162724
rs72345101
RepeatMasker

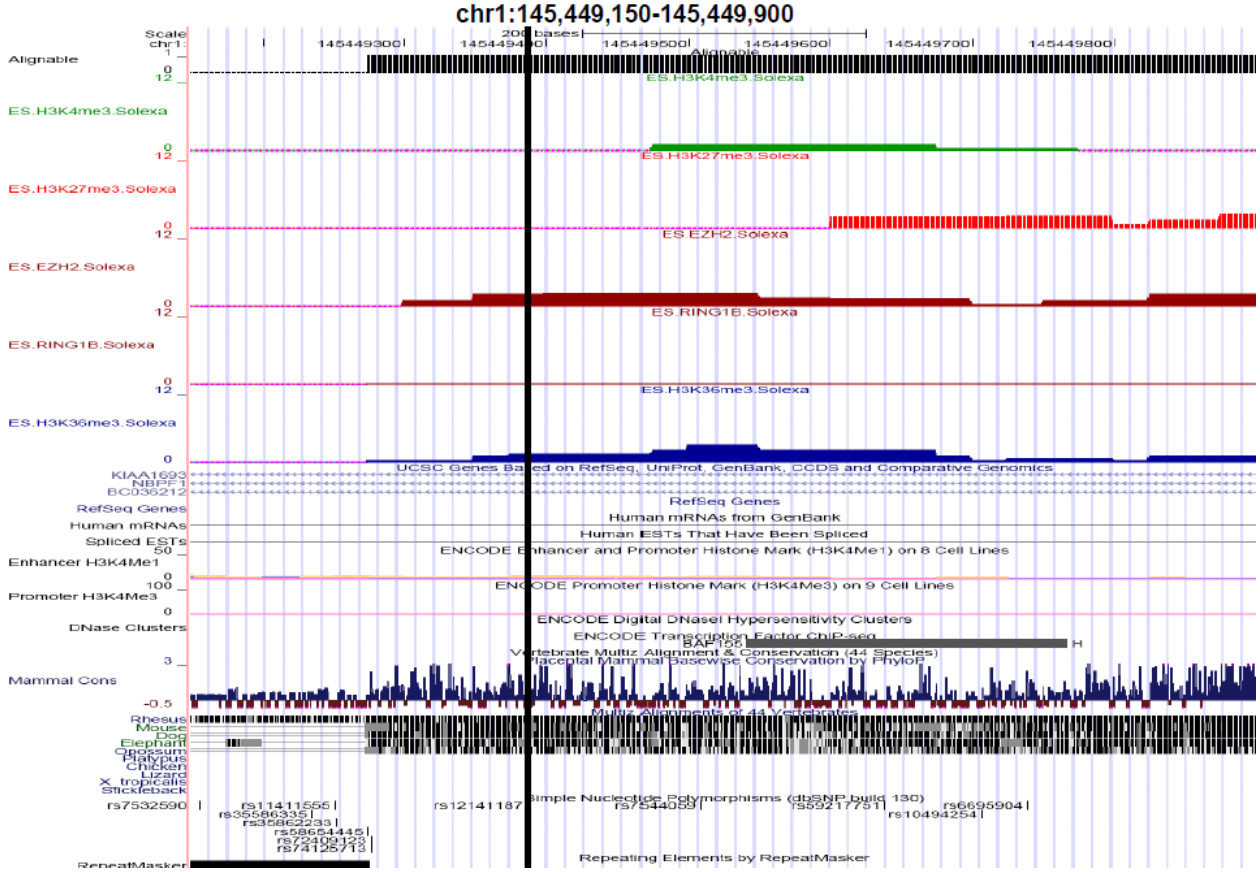
rs823128 Parkinson's Disease
size 526 bp.
chr1:203,979,725-203,980,250



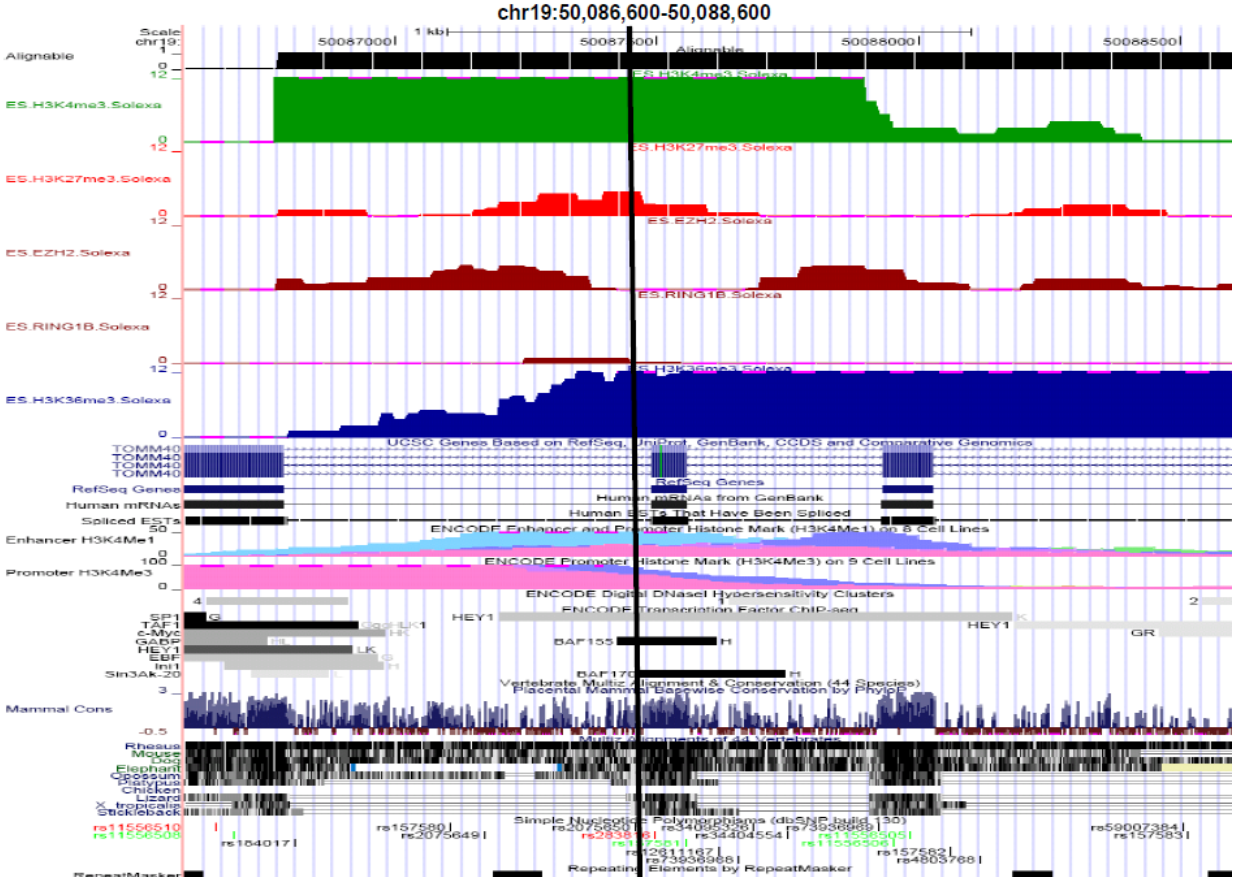
PD transRNA rs393152
size 676 bp.
chr17:41,074,575-41,075,250



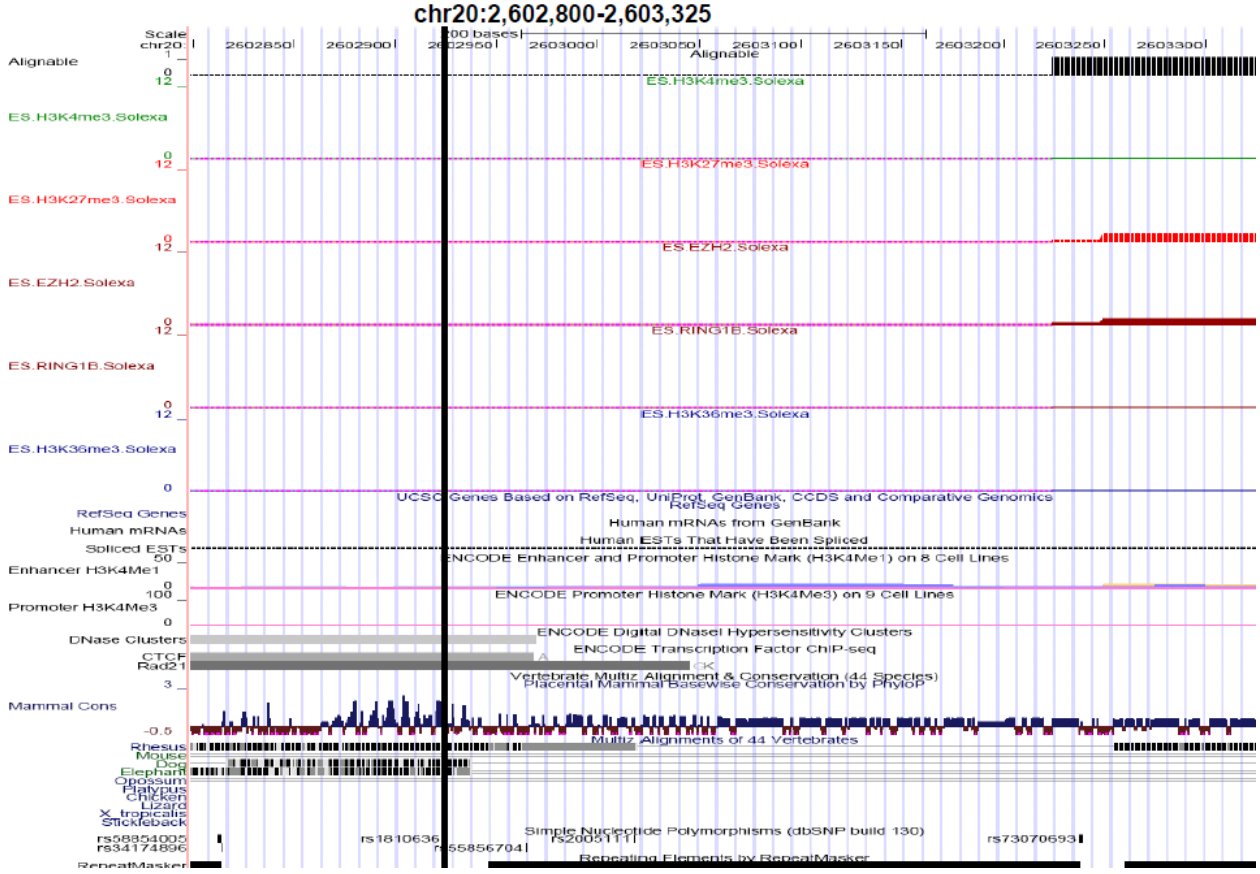
D26: rs12141187 SCHZ
size 751 bp.
chr1:145,449,150-145,449,900



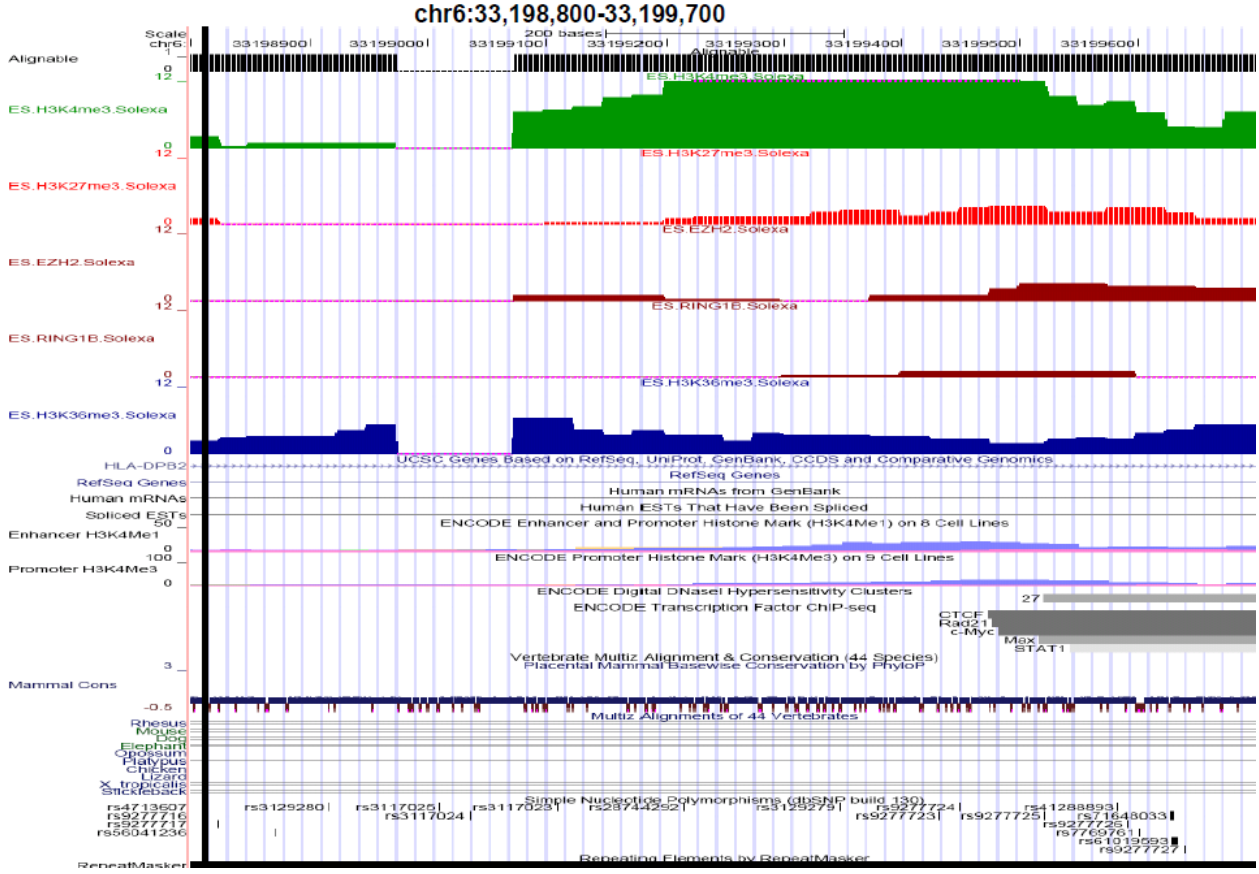
Longevity rs2075650
size 2,001 bp
chr19:50,086,600-50,088,600



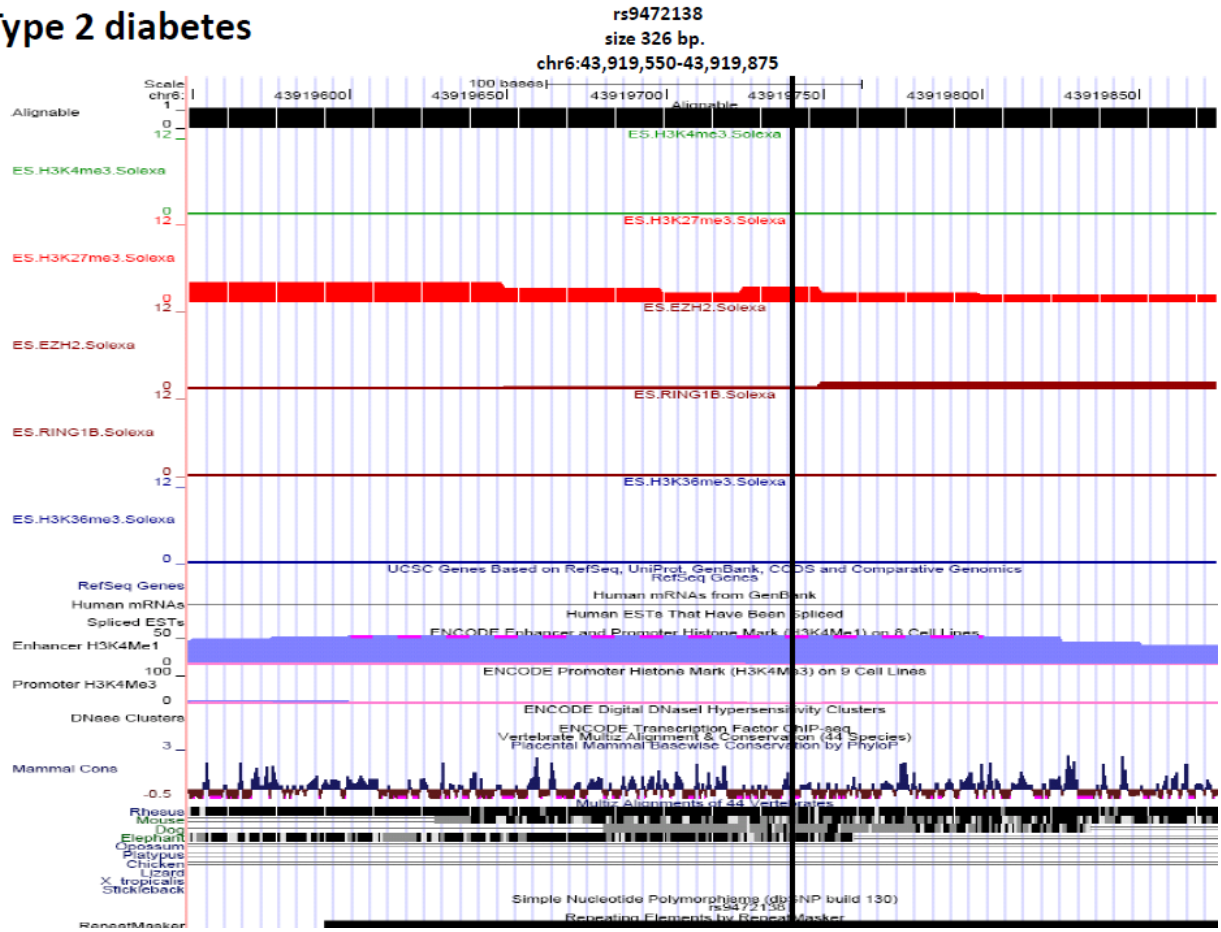
Longevity rs1810636
size 526 bp.
chr20:2,602,800-2,603,325



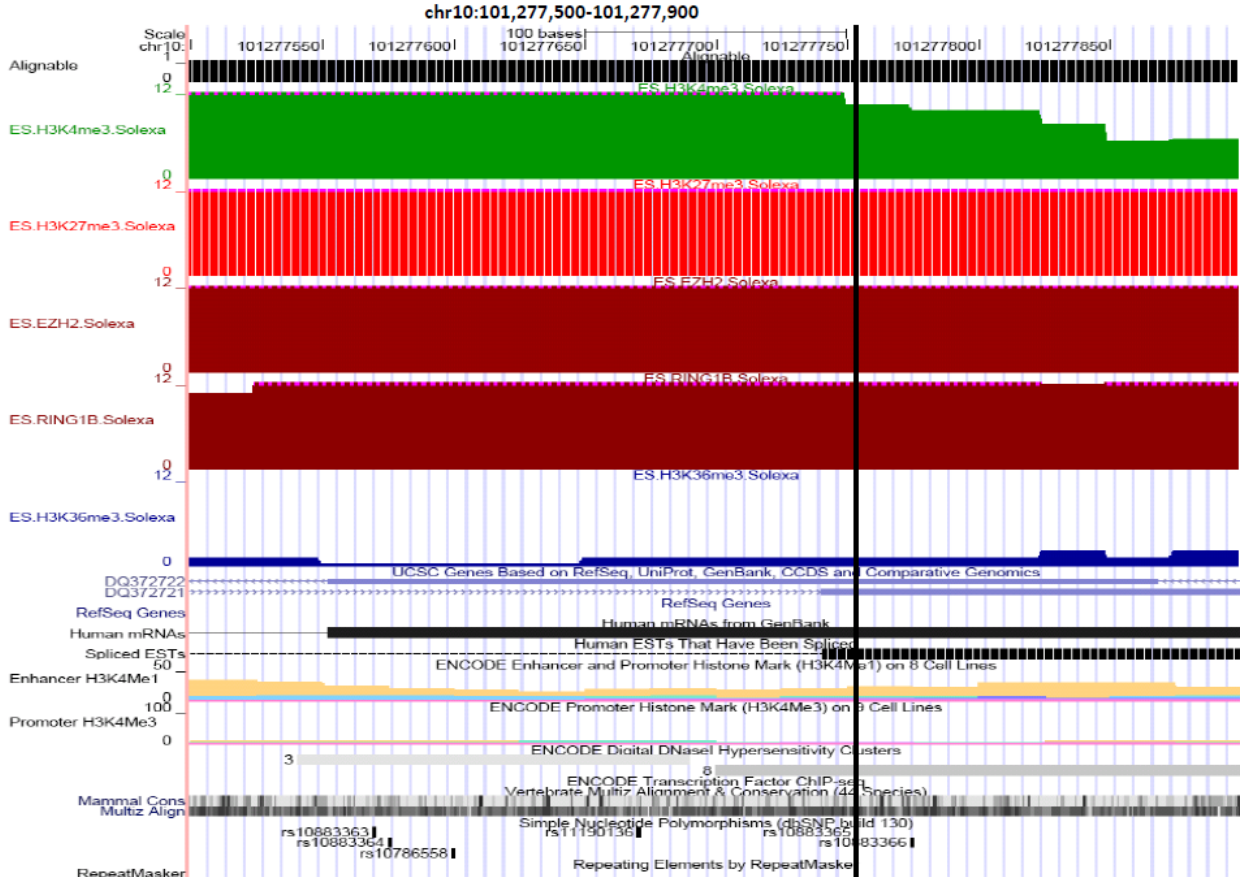
Longevity rs4713607
size 901 bp.
chr6:33,198,800-33,199,700



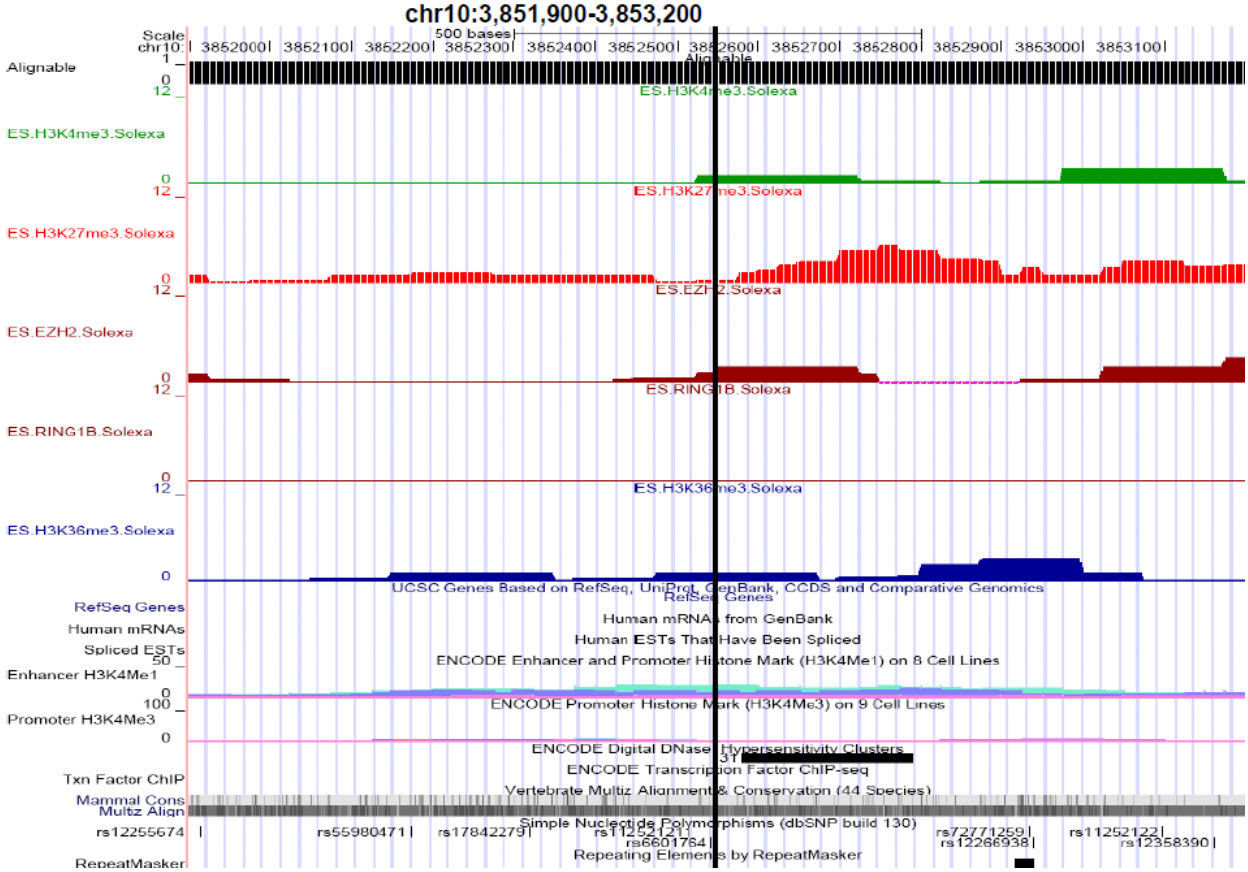
Type 2 diabetes



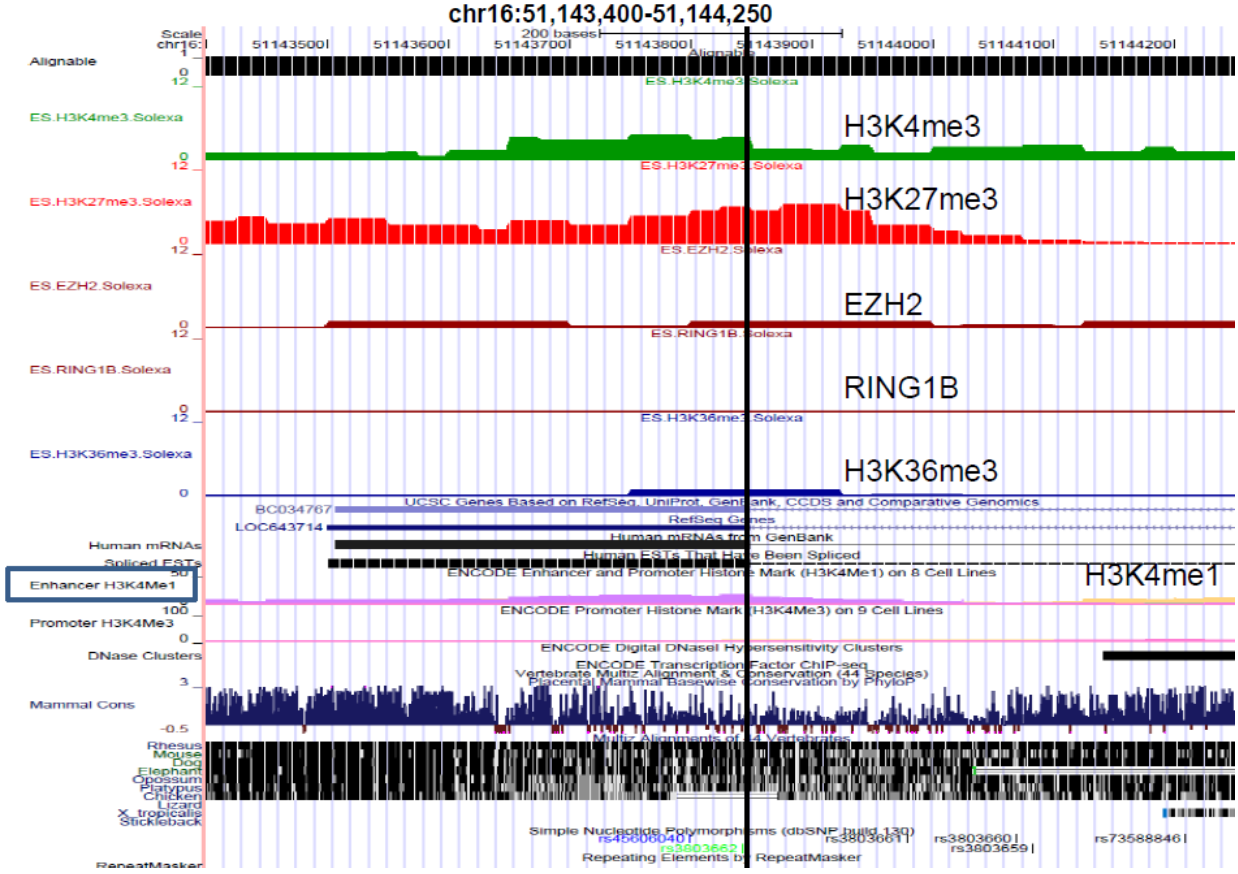
CD rs10883365
size 401 bp.
chr10:101,277,500-101,277,900



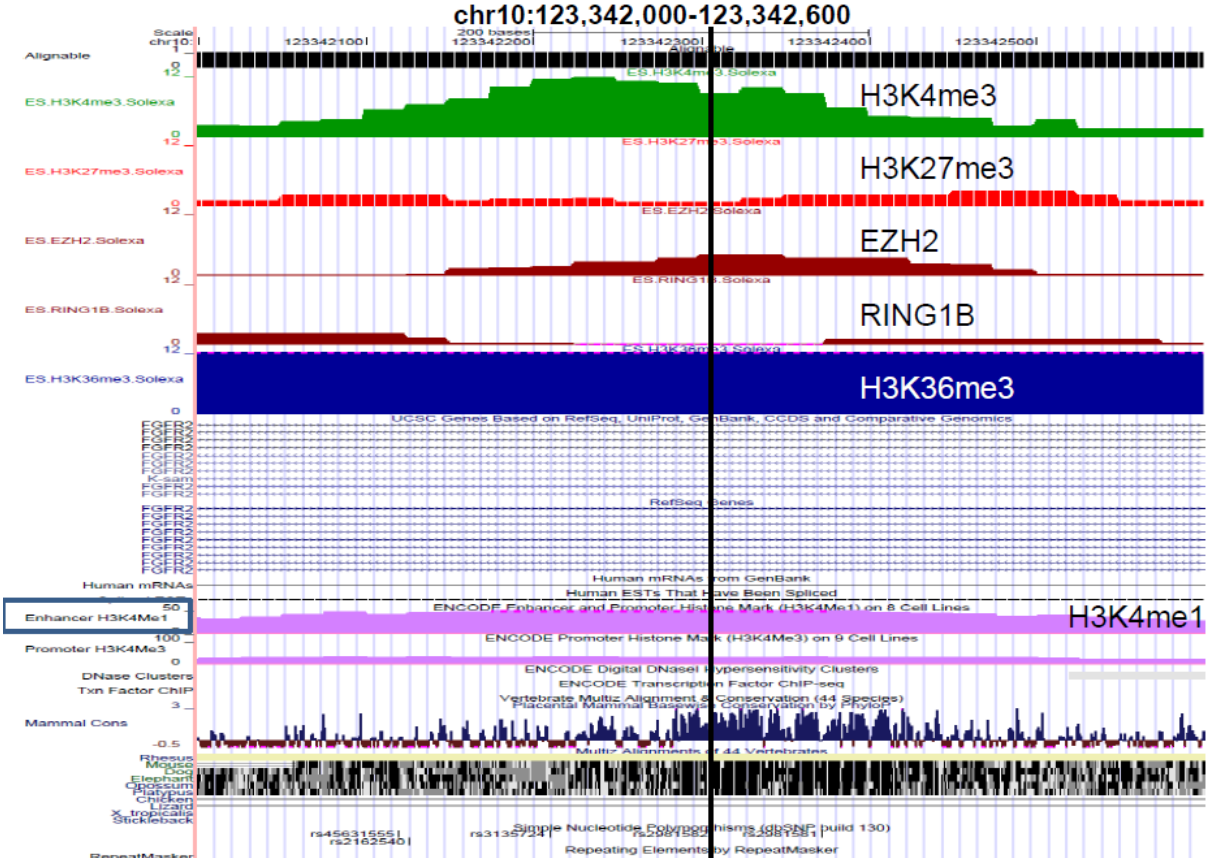
CD rs6601764
size 1,301 bp.
chr10:3,851,900-3,853,200

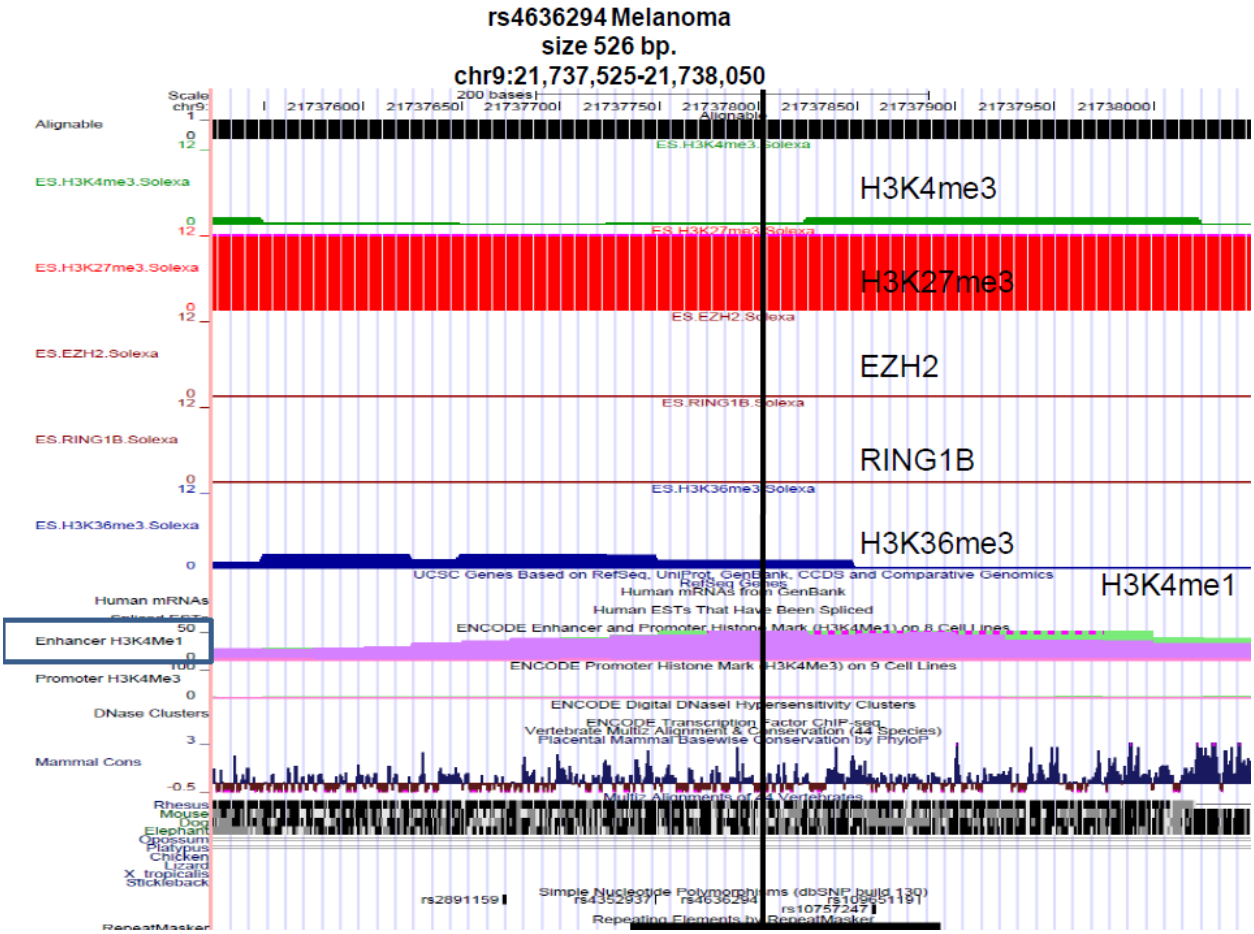


E3 rs3803662 Breast Cancer
size 851 bp.
chr16:51,143,400-51,144,250

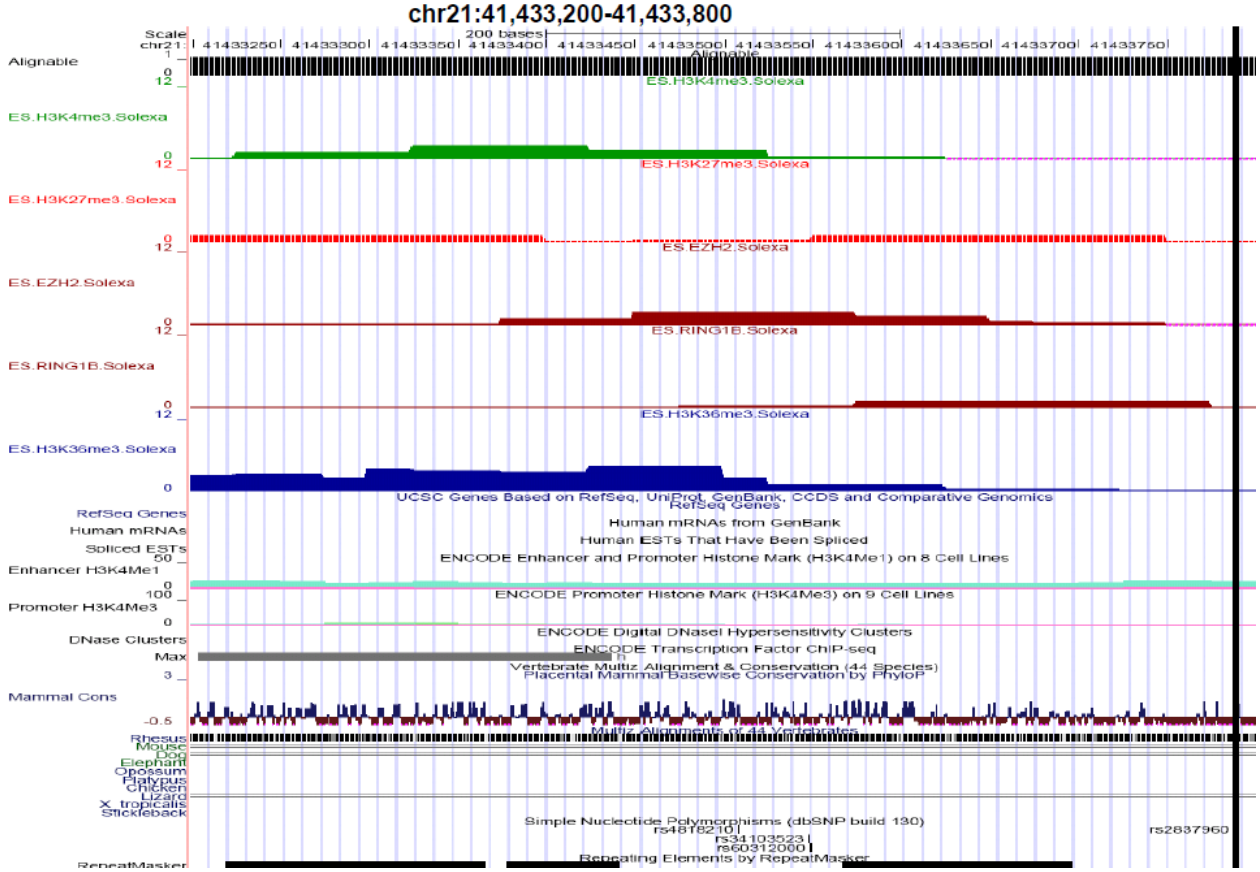


rs2981582 Breast Cancer Intronic FGFR2
size 601 bp.
chr10:123,342,000-123,342,600

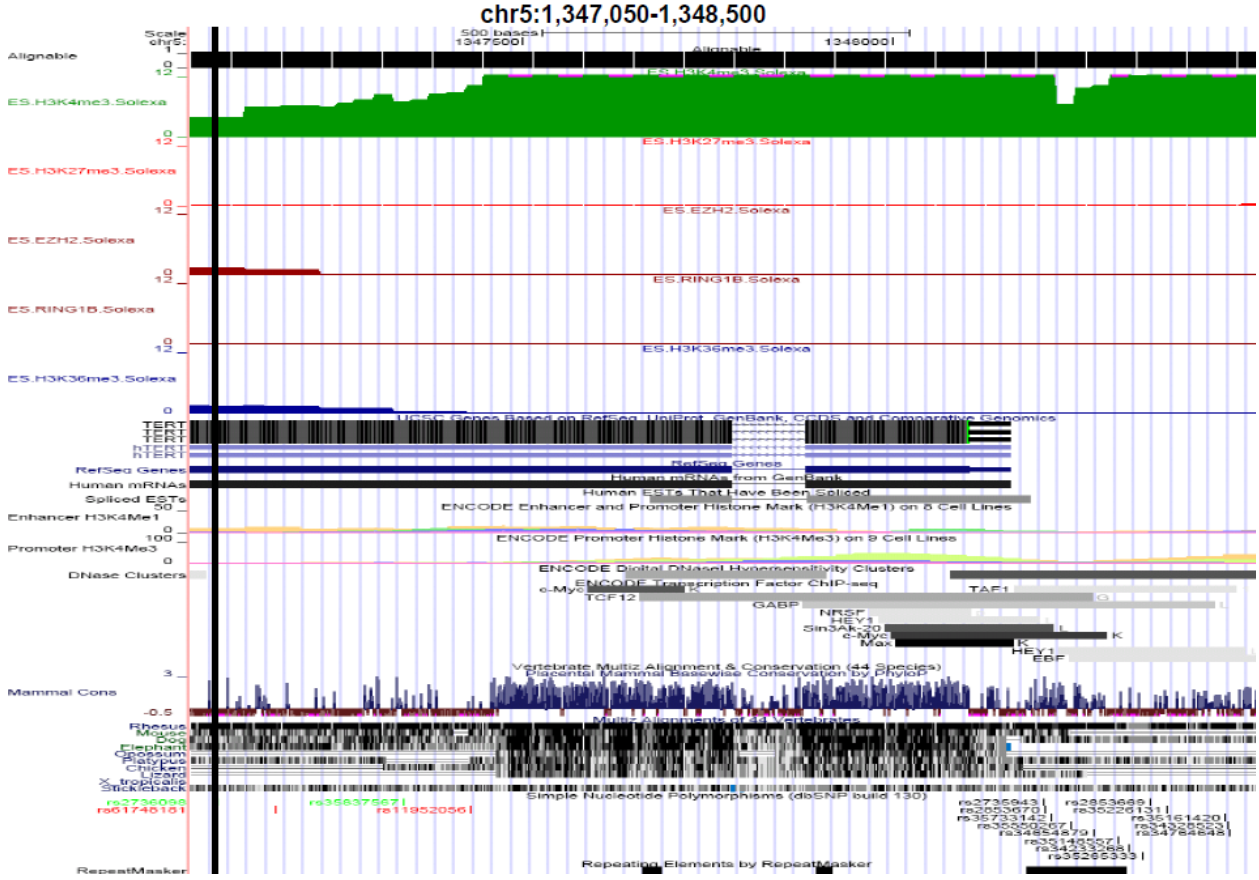



rs4636294 Melanoma
size 526 bp.
chr9:21,737,525-21,738,050
Alignable
ES.H3K4me3.Solexa
H3K4me3
ES.H3K27me3.Solexa
H3K27me3
ES.EZH2.Solexa
EZH2
ES.RING1B.Solexa
RING1B
ES.H3K36me3.Solexa
H3K36me3
H3K4me1
Human mRNAs
Enhancer H3K4Me1
Promoter H3K4Me3
DNase Clusters
Mammal Cons
Rhesus
Mouse
Dog
Elephant
Opossum
Platypus
Chicken
Lizard
X. tropicalis
Stickleback
rs2891159
rs4352937
rs4636294
rs10965119
rs10757247

RA D22 rs2837960
size 601 bp.
chr21:41,433,200-41,433,800



HT rs2736098
size 1,451 bp.
chr5:1,347,050-1,348,500



HDL rs7255436 ANGPTL4
size 10,251 bp.
chr19:8,335,000-8,345,250

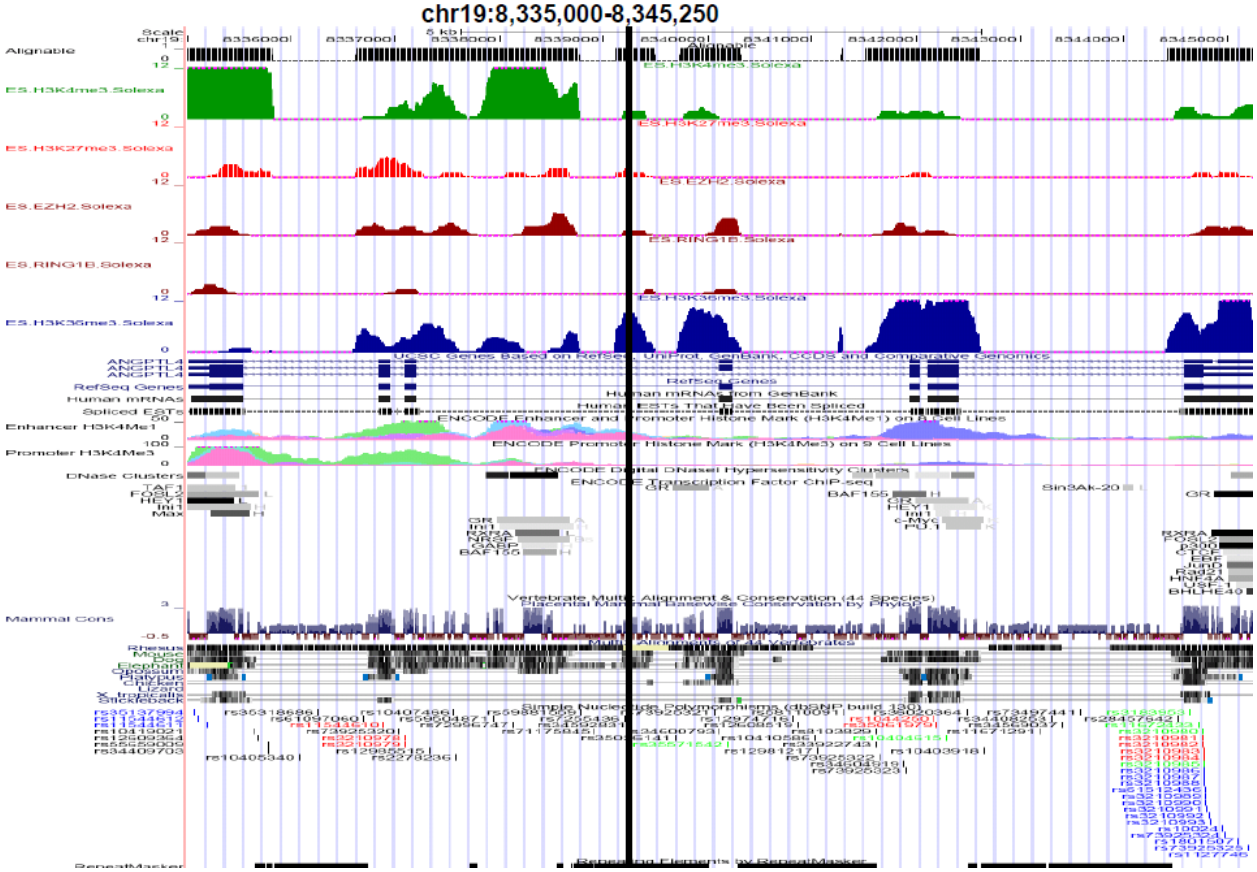


Figure S5.

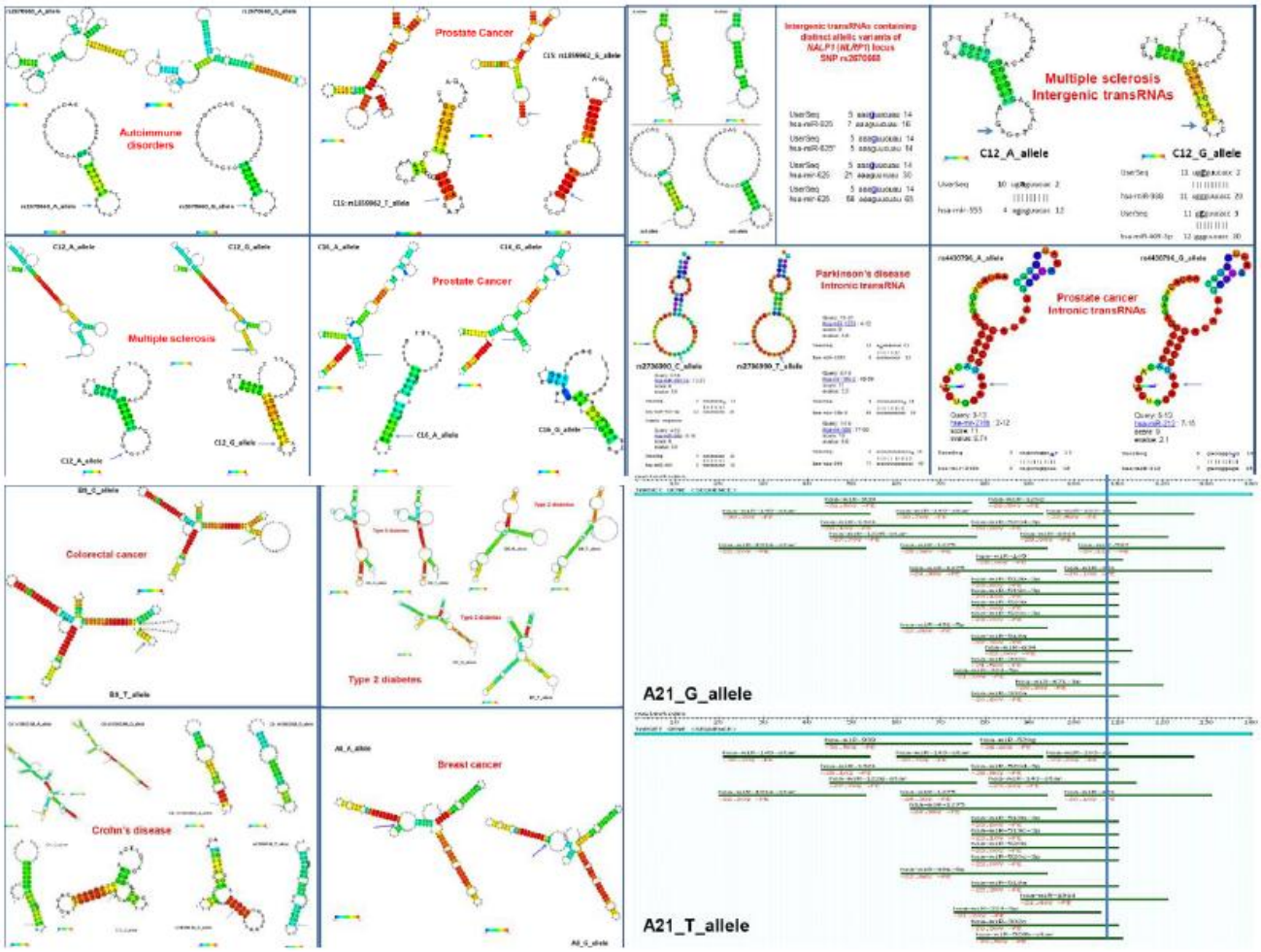
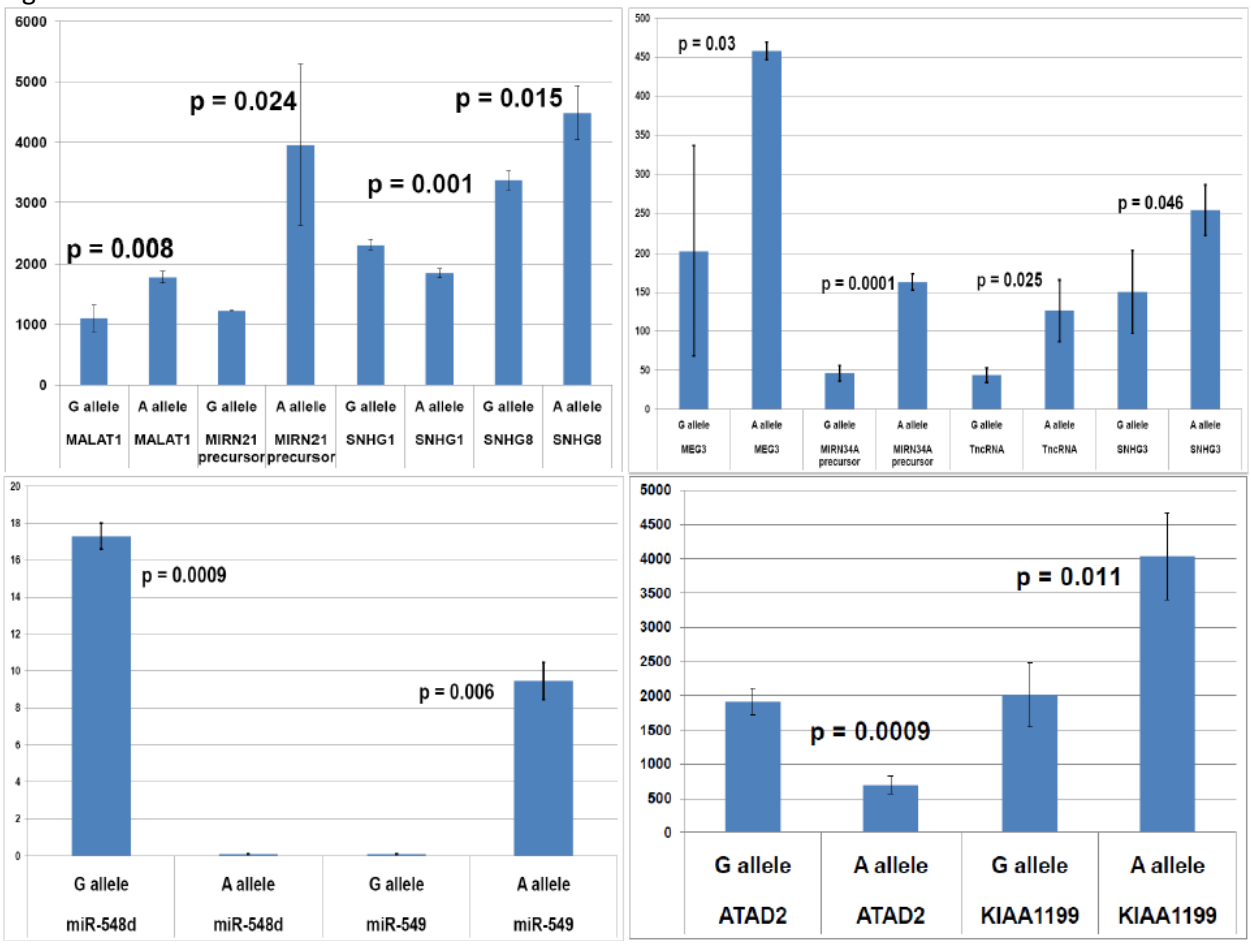
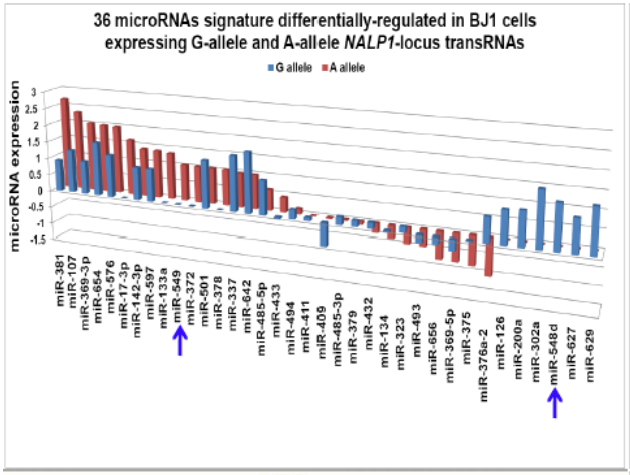
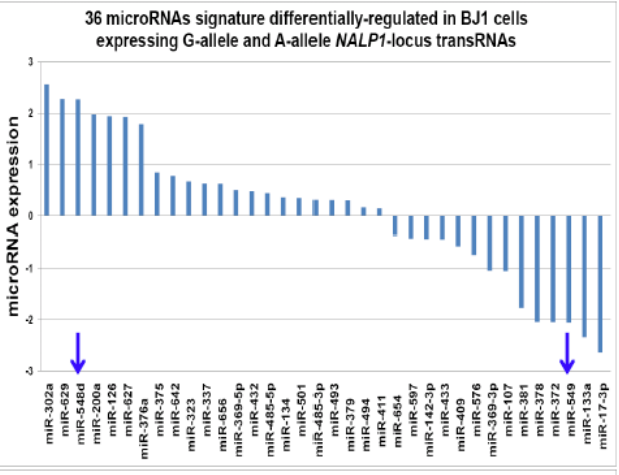
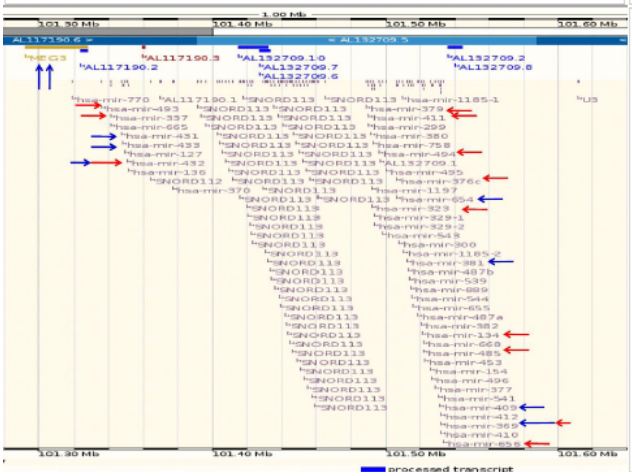

Autoimmune disorders
Prostate Cancer
Intergenic transRNAs containing distinct allelic variants of NALP1 (NLRP1) locus SNP rs2670660
Multiple sclerosis Intergenic transRNAs
C12_A_allele
C12_G_allele
Multiple sclerosis
Prostate Cancer
Parkinson's disease Intronic transRNA
Prostate cancer Intronic transRNAs
Colorectal cancer
Type 2 diabetes
A21_G_allele
Crohn's disease
Breast cancer
A21_T_allele

Figure S6.


p = 0.008
p = 0.024
p = 0.001
p = 0.015
G allele MALAT1
A allele MALAT1
G allele MIRN21 precursor
A allele MIRN21 precursor
G allele SNHG1
A allele SNHG1
G allele SNHG8
A allele SNHG8
p = 0.03
p = 0.0001
p = 0.025
p = 0.046
G allele MEG3
A allele MEG3
G allele MIRN34A precursor
A allele MIRN34A precursor
G allele TncRNA
A allele TncRNA
G allele SNHG3
A allele SNHG3
p = 0.0009
p = 0.006
G allele miR-548d
A allele miR-548d
G allele miR-549
A allele miR-549
p = 0.0009
p = 0.011
G allele ATAD2
A allele ATAD2
G allele KIAA1199
A allele KIAA1199


36 microRNAs signature differentially-regulated in BJ1 cells expressing G-allele and A-allele NALP1-locus transRNAs
G allele
A allele
microRNA expression
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
miR-381
miR-107
miR-369-3p
miR-654
miR-576
miR-17-3p
miR-142-3p
miR-597
miR-133a
miR-549
miR-372
miR-501
miR-378
miR-337
miR-642
miR-485-5p
miR-433
miR-494
miR-411
miR-409
miR-485-3p
miR-379
miR-432
miR-134
miR-323
miR-493
miR-656
miR-369-5p
miR-375
miR-376a-2
miR-126
miR-200a
miR-302a
miR-548d
miR-627
miR-629


36 microRNAs signature differentially-regulated in BJ1 cells expressing G-allele and A-allele NALP1-locus transRNAs
microRNA expression
3
2
1
0
-1
-2
-3
miR-302a
miR-629
miR-548d
miR-200a
miR-126
miR-627
miR-376a
miR-375
miR-642
miR-323
miR-337
miR-656
miR-369-5p
miR-432
miR-485-5p
miR-134
miR-501
miR-485-3p
miR-493
miR-379
miR-494
miR-411
miR-654
miR-597
miR-142-3p
miR-433
miR-409
miR-576
miR-369-3p
miR-107
miR-381
miR-378
miR-372
miR-549
miR-133a
miR-17-3p


1.00 Mb
101.30 Mb
101.40 Mb
101.50 Mb
101.60 Mb
AL117190.6
AL132709.5
MEG3
AL117190.3
AL132709.1.0
AL132709.2
AL117190.2
AL132709.7
AL132709.8
AL132709.6
hsa-mir-770
AL117190.1
SNORD113
hsa-mir-1185-1
hsa-mir-493
hsa-mir-379
hsa-mir-337
hsa-mir-411
hsa-mir-665
hsa-mir-299
hsa-mir-431
hsa-mir-380
hsa-mir-433
hsa-mir-758
hsa-mir-127
hsa-mir-494
hsa-mir-432
AL132709.1
hsa-mir-136
hsa-mir-495
hsa-mir-376c
hsa-mir-376
hsa-mir-1197
hsa-mir-654
hsa-mir-329
hsa-mir-329-1
hsa-mir-329-2
hsa-mir-543
hsa-mir-300
hsa-mir-1185-2
hsa-mir-381
hsa-mir-487b
hsa-mir-539
hsa-mir-889
hsa-mir-544
hsa-mir-655
hsa-mir-487a
hsa-mir-382
hsa-mir-134
hsa-mir-668
hsa-mir-485
hsa-mir-453
hsa-mir-154
hsa-mir-496
hsa-mir-377
hsa-mir-541
hsa-mir-409
hsa-mir-412
hsa-mir-369
hsa-mir-410
hsa-mir-655
processed transcript

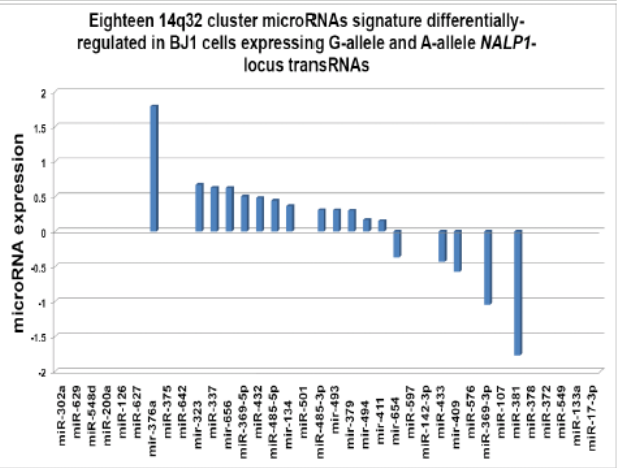

Eighteen 14q32 cluster microRNAs signature differentially-regulated in BJ1 cells expressing G-allele and A-allele NALP1-locus transRNAs
microRNA expression
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
miR-302a
miR-629
miR-548d
miR-200a
miR-126
miR-627
mir-376a
miR-375
miR-642
mir-323
miR-337
mir-656
miR-369-5p
miR-432
miR-485-5p
mir-134
miR-501
miR-485-3p
miR-493
mir-379
mir-494
mir-411
mir-654
miR-597
miR-142-3p
miR-433
mir-409
miR-576
miR-369-3p
miR-107
miR-381
miR-378
miR-372
miR-549
miR-133a
miR-17-3p

Figure S7.

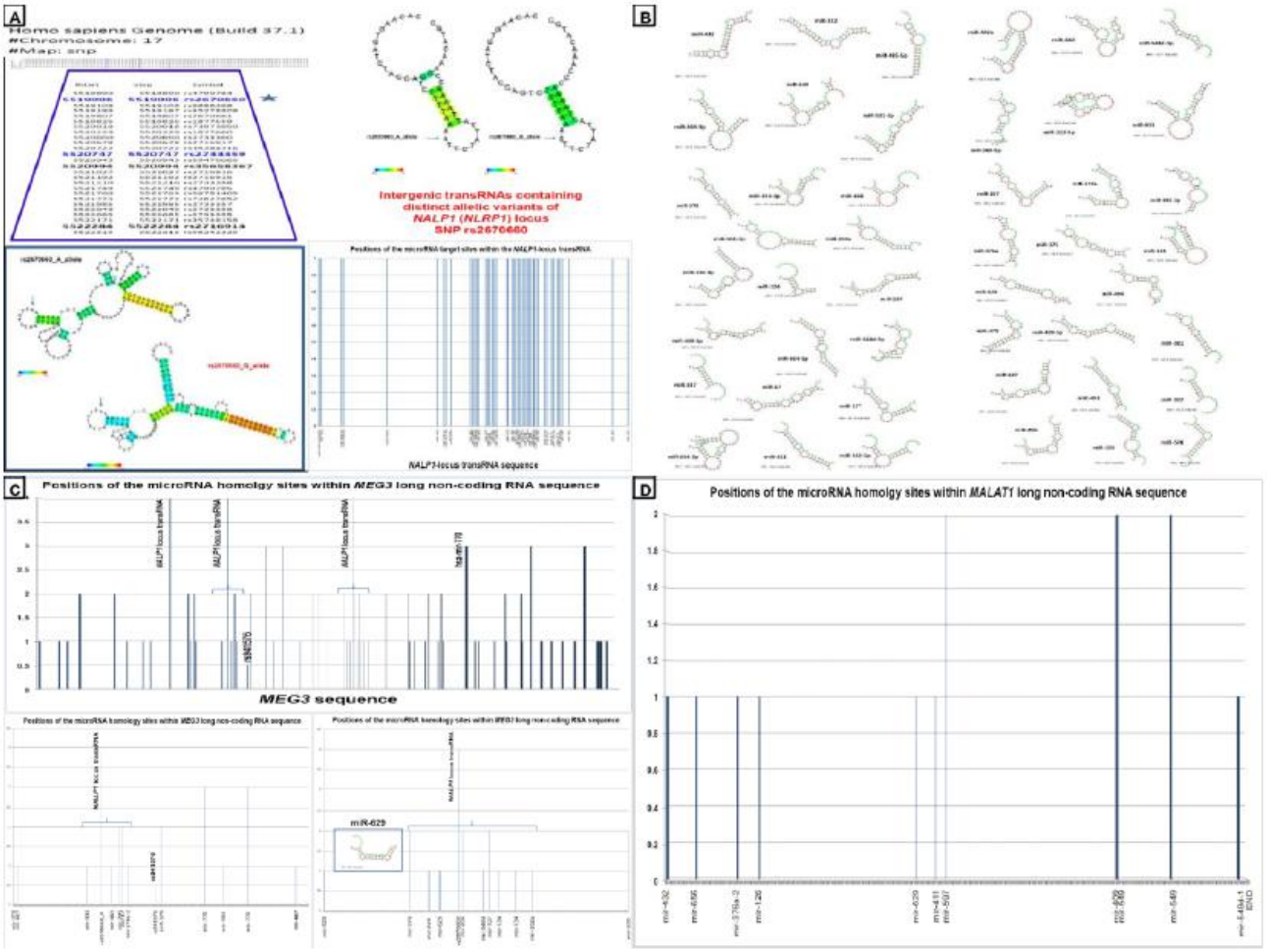
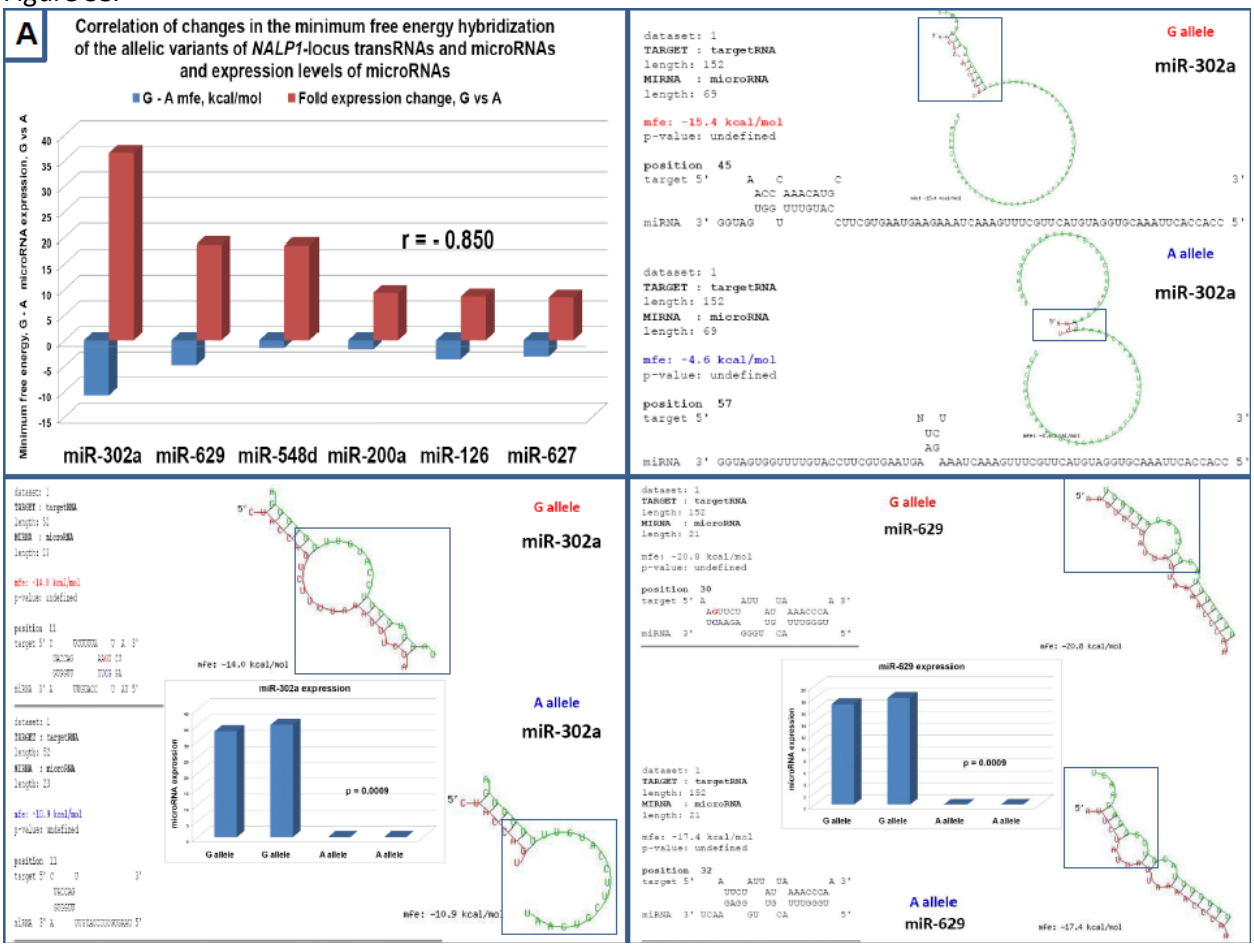
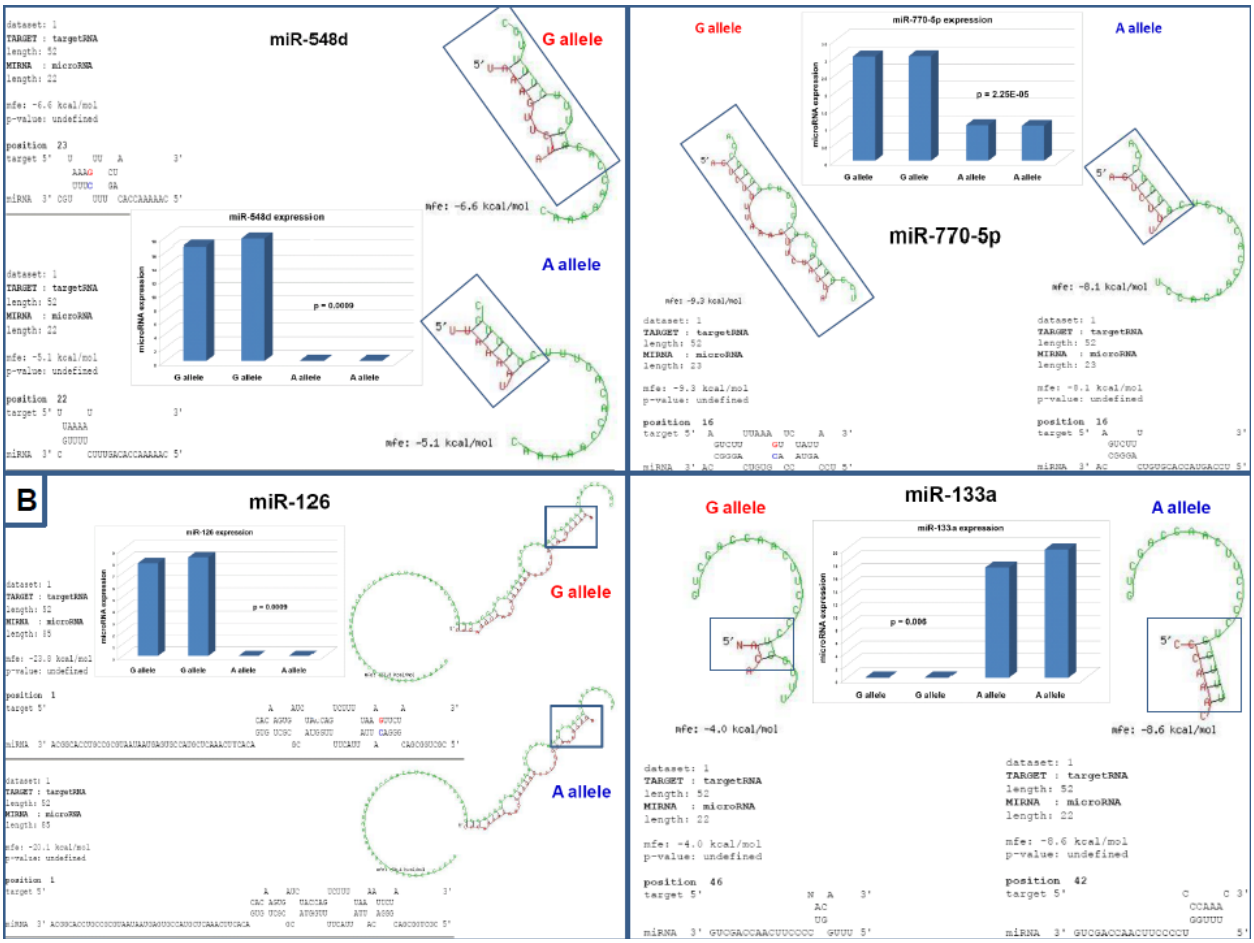
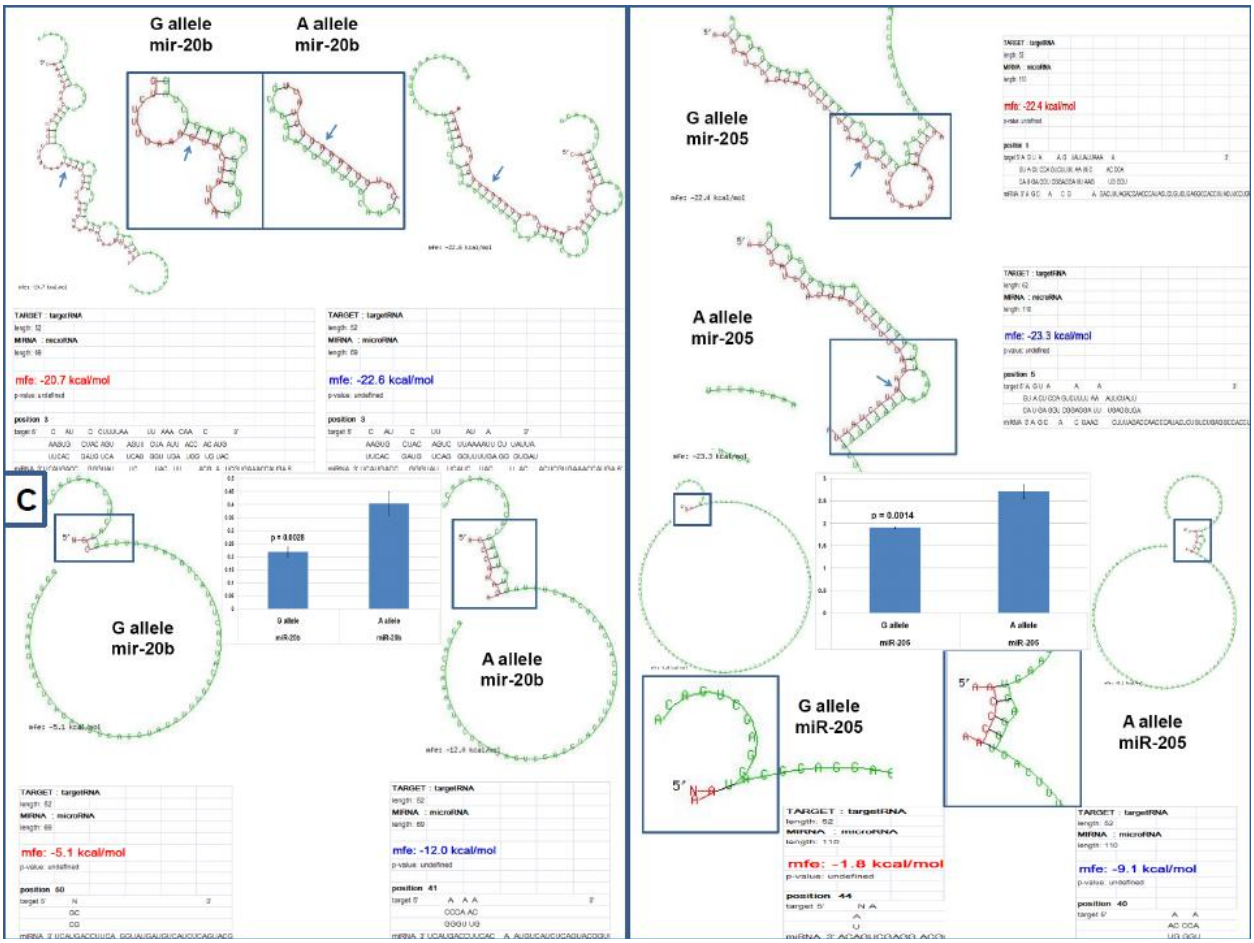
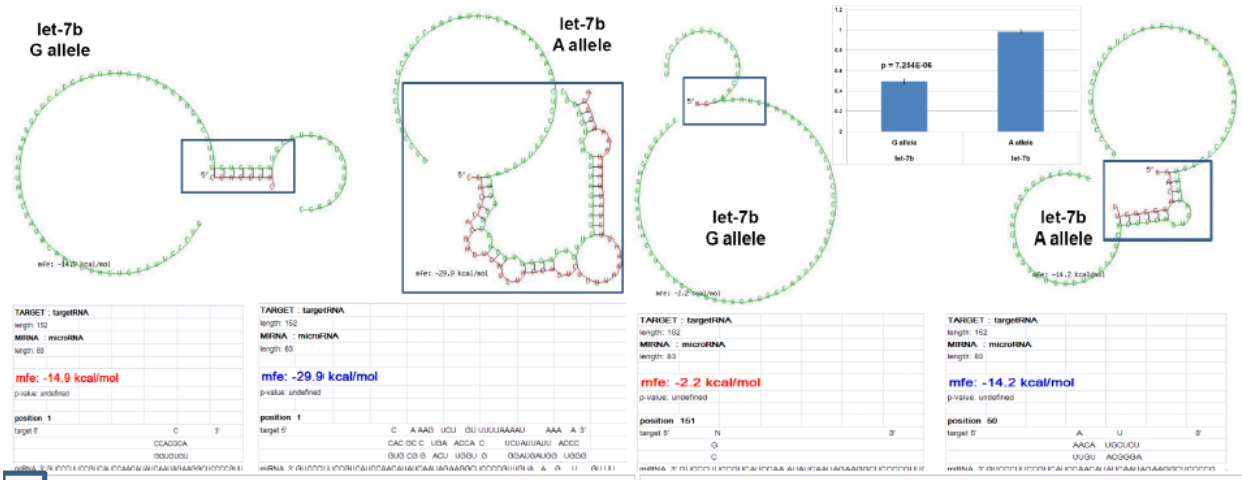

A
Homo sapiens Genome (Build 37.1)
#Chromosome: 17
#Map: snp
Intergenic transRNAs containing distinct allelic variants of NALP1 (NLRP1) locus SNP rs2670660
NALP1-locus transRNA sequence
B
C
Positions of the microRNA homology sites within MEG3 long non-coding RNA sequence
MEG3 sequence
miR-625
D
Positions of the microRNA homology sites within MALAT1 long non-coding RNA sequence

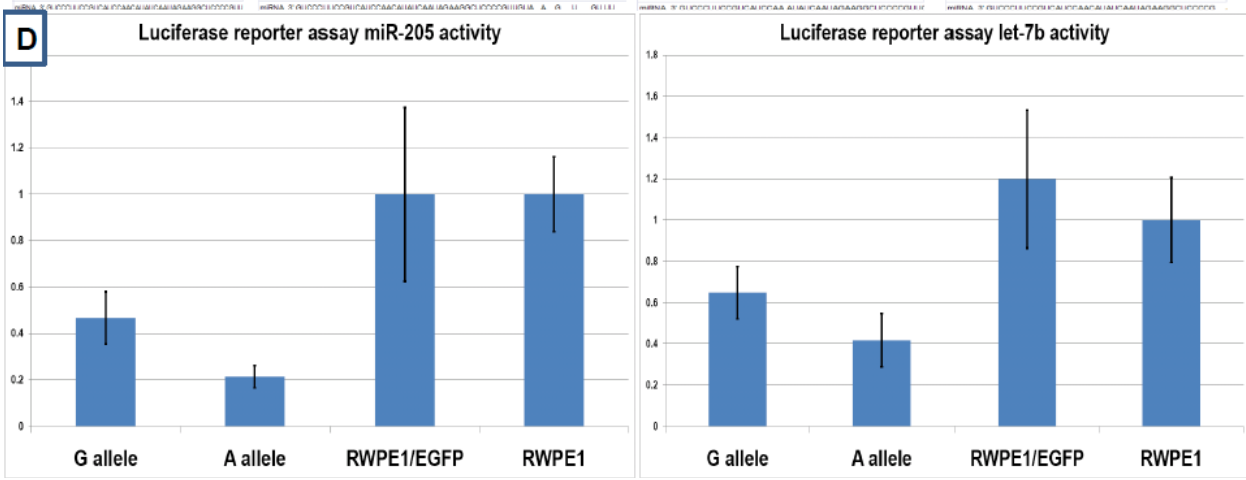
Figure S8.


A
Correlation of changes in the minimum free energy hybridization of the allelic variants of NALP1-locus transRNAs and microRNAs and expression levels of microRNAs
G - A mfe, kcal/mol
Fold expression change, G vs A
Minimum free energy, G - A microRNA expression, G vs A
r = - 0.850
miR-302a
miR-629
miR-548d
miR-200a
miR-126
miR-627
G allele
miR-302a
A allele
miR-302a
G allele
miR-629
A allele
miR-629
miR-302a expression
miR-629 expression
p = 0.0009
G allele
G allele
A allele
A allele


miR-548d
G allele
A allele
miR-548d expression
p = 0.0009
G allele
G allele
A allele
A allele
mfe: -6.6 kcal/mol
mfe: -5.1 kcal/mol
miR-770-5p
miR-770-5p expression
p = 2.25E-05
mfe: -9.3 kcal/mol
mfe: -8.1 kcal/mol
B
miR-126
miR-126 expression
p = 0.0009
mfe: -23.0 kcal/mol
mfe: -20.1 kcal/mol
miR-133a
miR-133a expression
p = 0.006
mfe: -4.0 kcal/mol
mfe: -8.6 kcal/mol


G allele
mir-20b
A allele
mir-20b
mfe: -20.7 kcal/mol
mfe: -22.6 kcal/mol
C
G allele
mir-20b
A allele
mir-20b
p = 0.0028
mfe: -5.1 kcal/mol
mfe: -12.0 kcal/mol
G allele
mir-205
A allele
mir-205
mfe: -22.4 kcal/mol
mfe: -23.3 kcal/mol
p = 0.0014
G allele
miR-205
A allele
miR-205
mfe: -1.8 kcal/mol
mfe: -9.1 kcal/mol


let-7b
G allele
let-7b
A allele
p = 7.234E-06
G allele
A allele
let-7b
let-7b
G allele
let-7b
A allele
mfe: -14.9 kcal/mol
mfe: -29.9 kcal/mol
mfe: -2.2 kcal/mol
mfe: -14.2 kcal/mol


D
Luciferase reporter assay miR-205 activity
1.4
1.2
1
0.8
0.6
0.4
0.2
0
G allele
A allele
RWPE1/EGFP
RWPE1
Luciferase reporter assay let-7b activity
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
G allele
A allele
RWPE1/EGFP
RWPE1

E **Allele equilibrium model of mechanisms of action of *NALP1*-locus trans-RNAs**

CCACGCACAAGTGATCTACCAGTCTTTTAAA[G/A]TTCTATTATTAAAACCCAAACATGCTCTTTCATTTCCACAGAACACTGGGTCTAAATTTAGACTGGTGCATCCTGATGCTGCACCAGTCTGCTCTTAATTTAAGGCATACAGGCATCTTG

CACAAGTGATCTACCAGTCTTTTAAA[G/A]TTCTATTATTAAAACCCAAACATGC

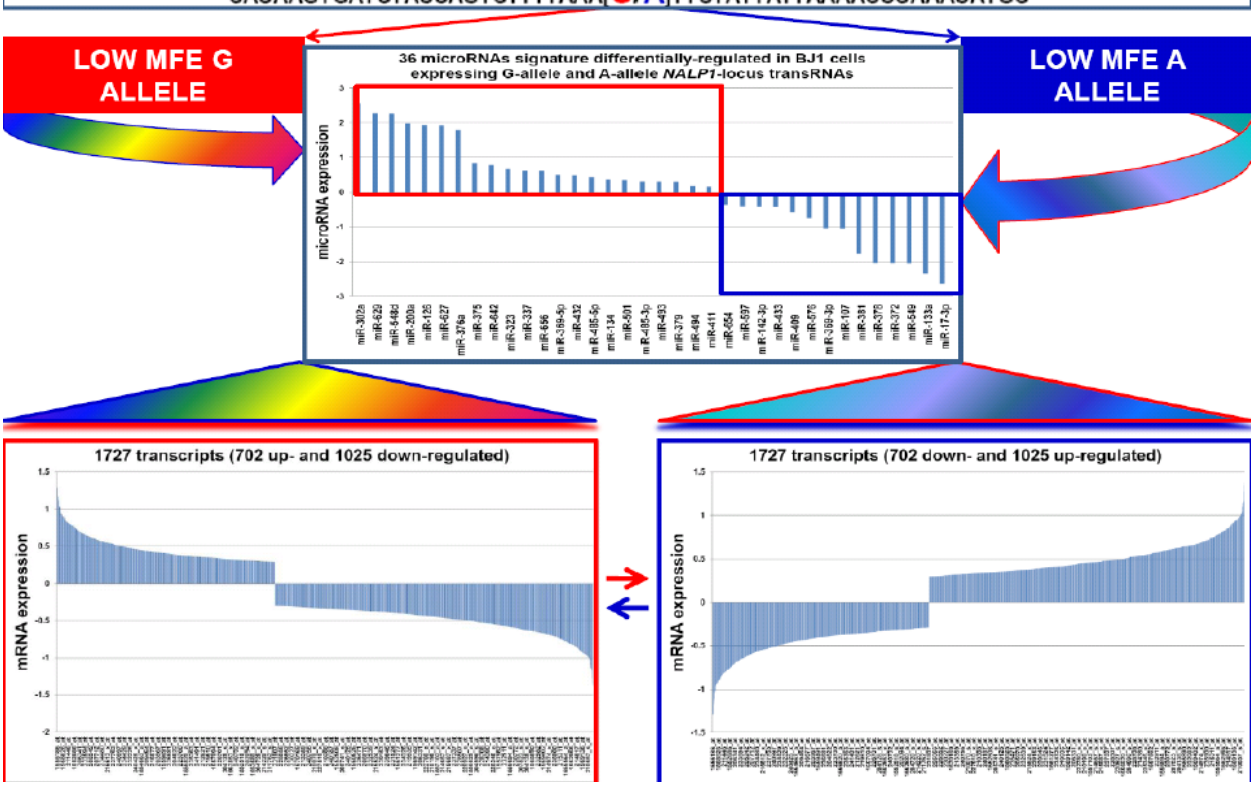
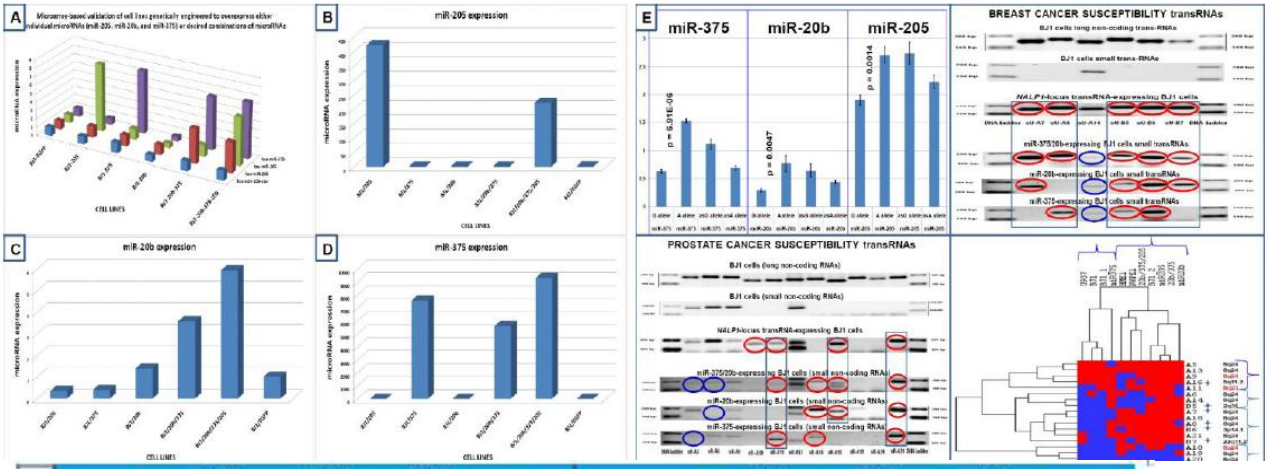
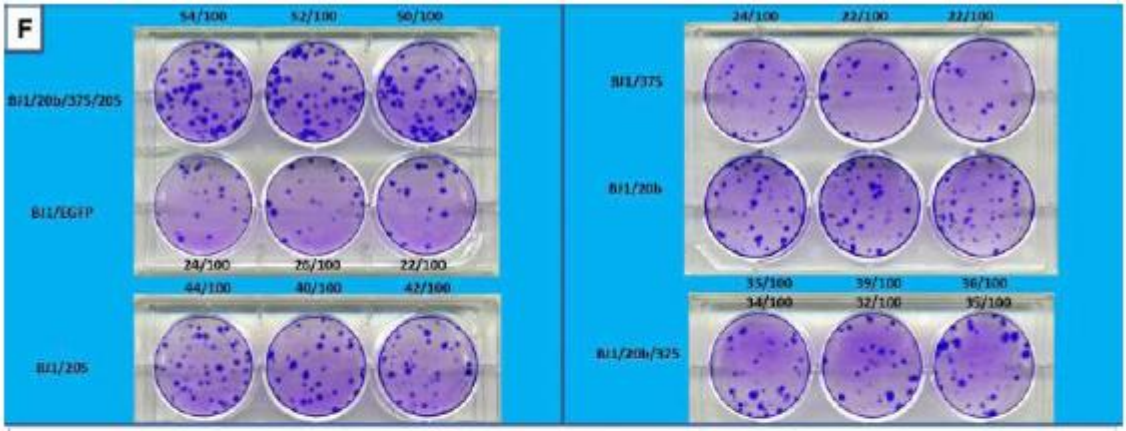


Figure S9.


A
B
miR-205 expression
E
miR-375
miR-20b
miR-205
p = 6.01E-06
p = 0.0047
p = 0.0014
BREAST CANCER SUSCEPTIBILITY transRNAs
C
miR-20b expression
D
miR-375 expression
PROSTATE CANCER SUSCEPTIBILITY transRNAs


F
54/100
52/100
50/100
BJ1/20b/375/205
BJ1/EGFP
24/100
26/100
22/100
44/100
40/100
42/100
BJ1/205
24/100
22/100
22/100
BJ1/375
BJ1/20b
35/100
39/100
36/100
34/100
32/100
35/100
BJ1/20b/325

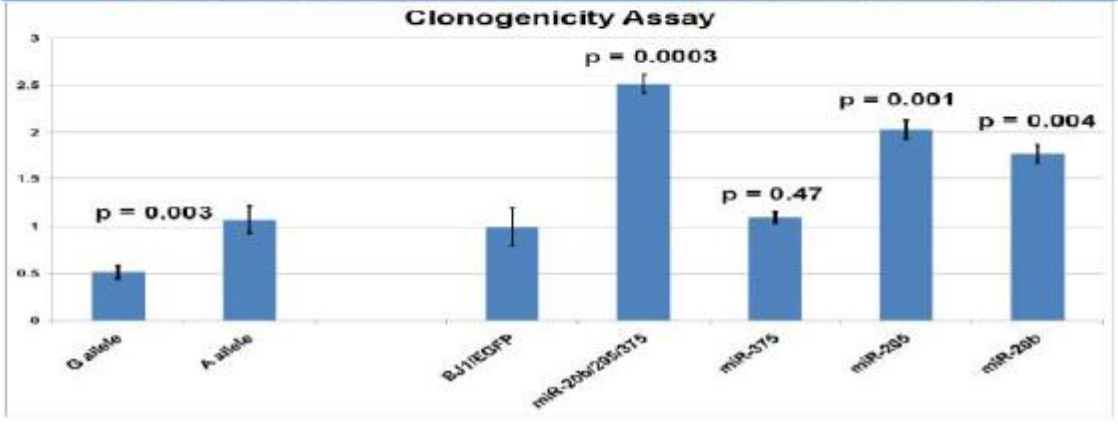
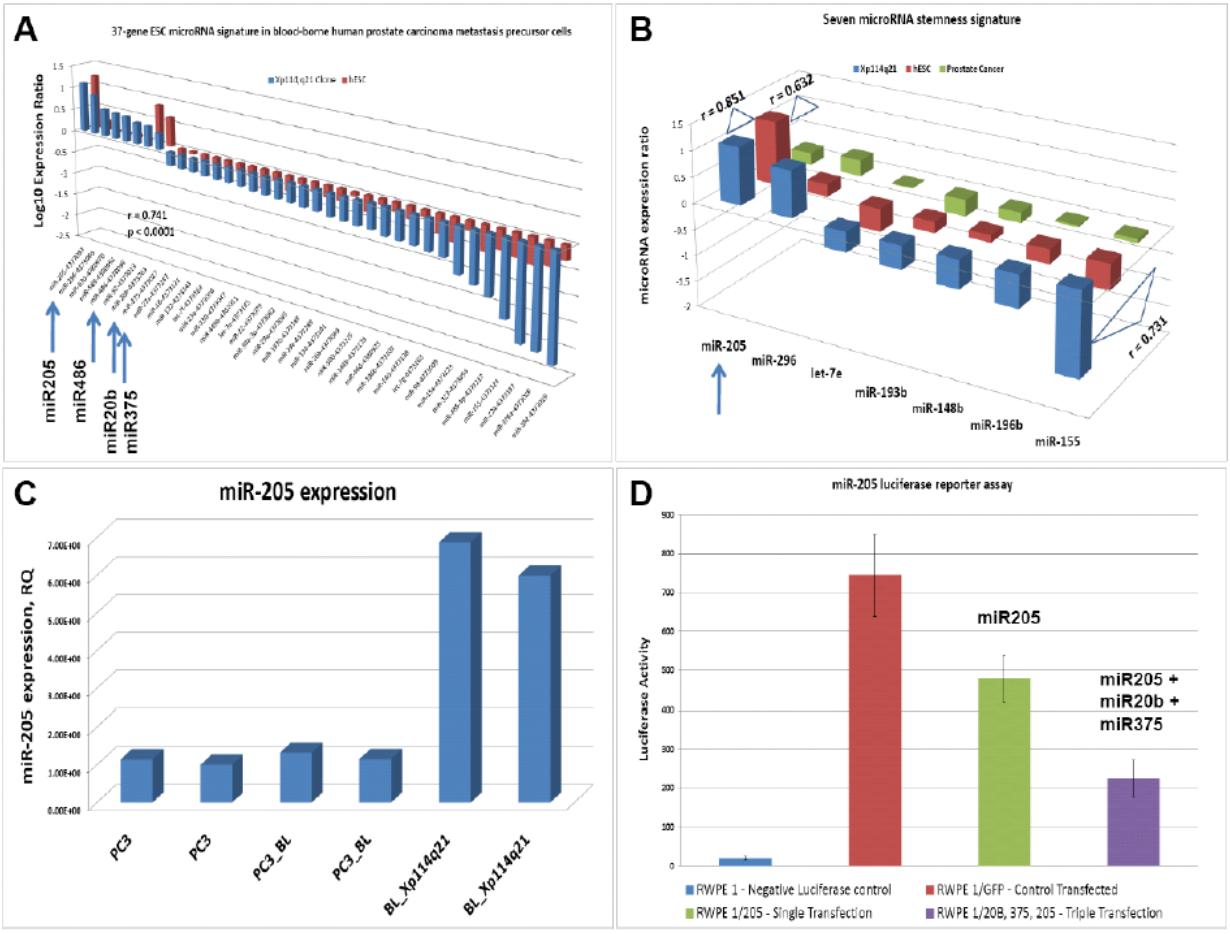
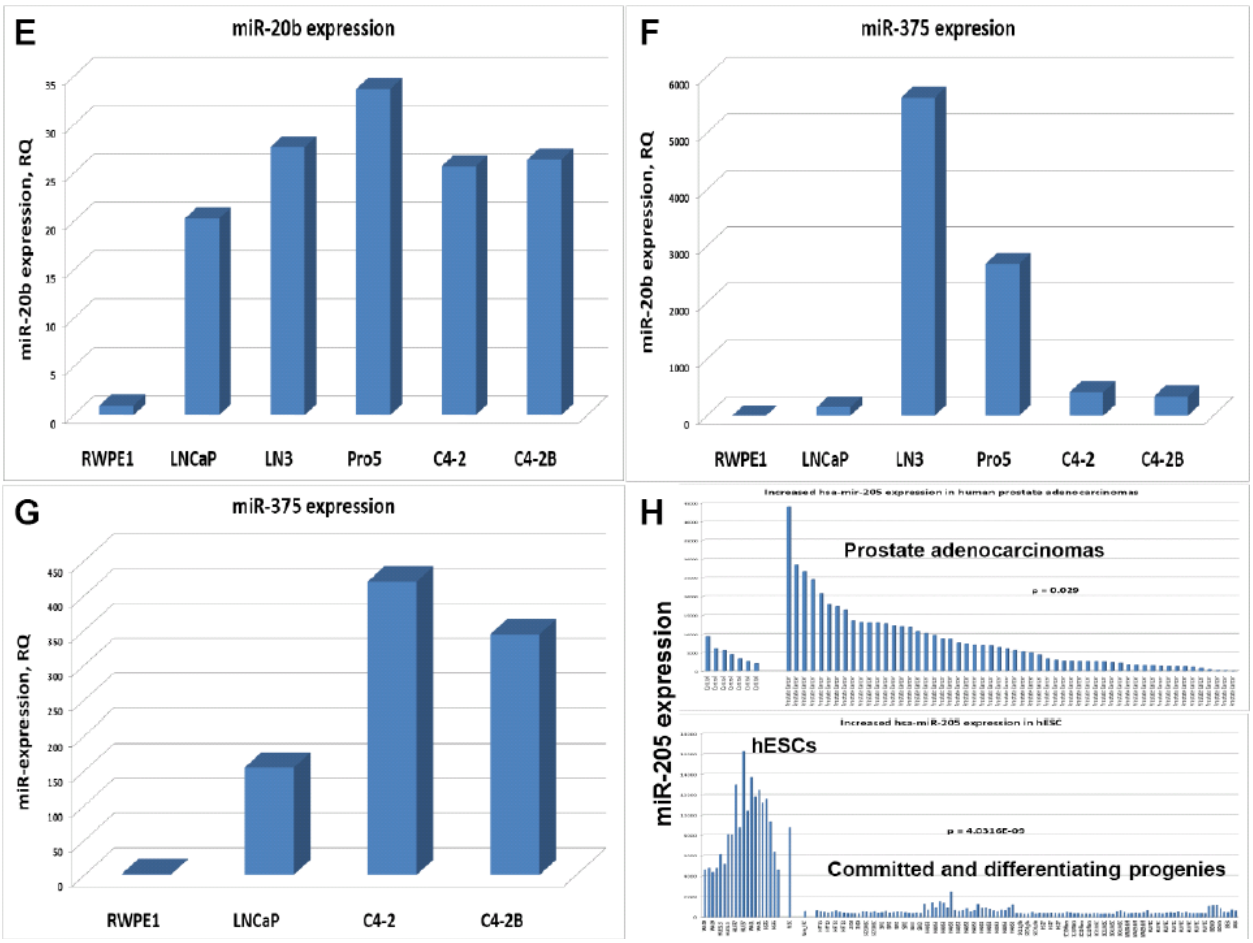

Clonogenicity Assay
p = 0.0003
p = 0.001
p = 0.004
p = 0.003
p = 0.47
G allele
A allele
BJ1/EGFP
miR-20b/205/375
miR-375
miR-205
miR-20b

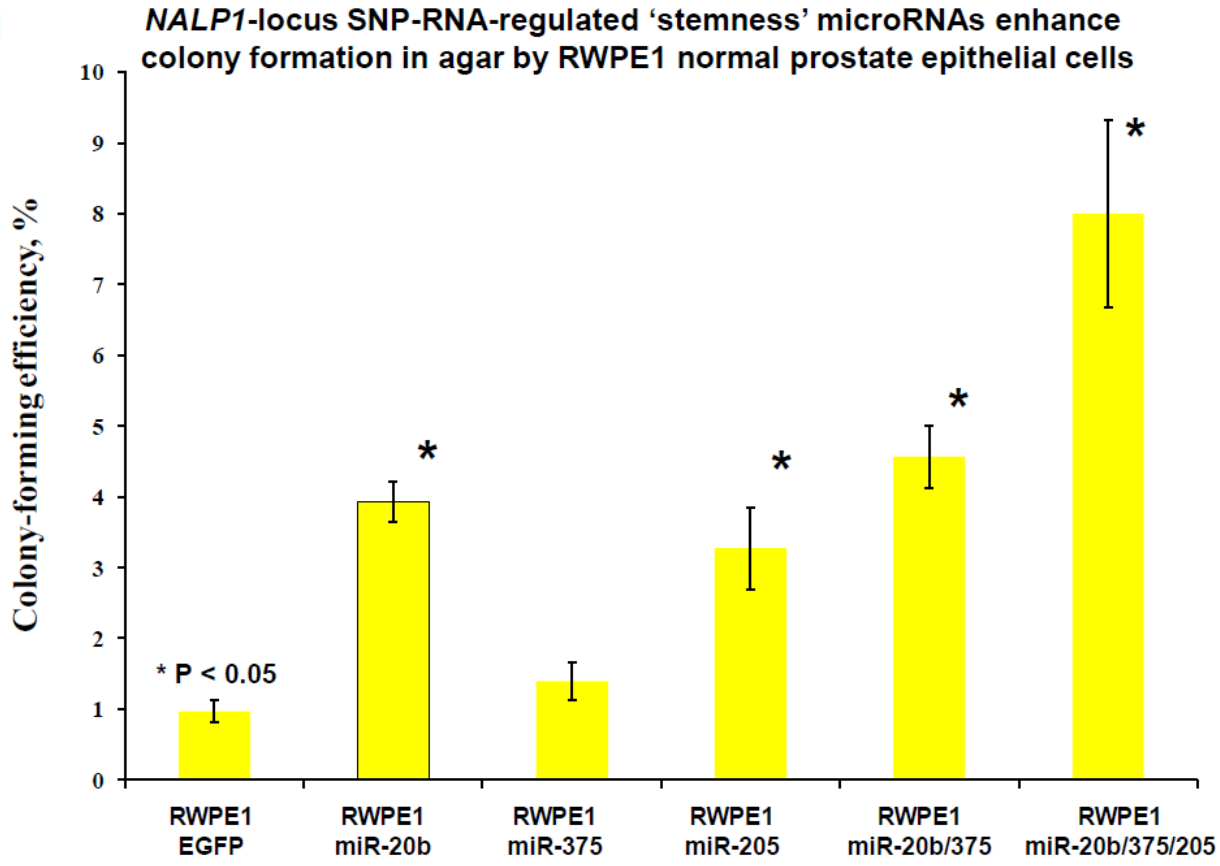
Figure S10.

Expression of NALP1-locus SNP-RNA-regulated “stemness” microRNA in human prostate cancer


A
37-gene ESC microRNA signature in blood-borne human prostate carcinoma metastasis precursor cells
Log10 Expression Ratio
Xp114q21 Clone
hESC
r = 0.741
p < 0.0001
miR205
miR486
miR20b
miR375
B
Seven microRNA stemness signature
Xp114q21
hESC
Prostate Cancer
microRNA expression ratio
r = 0.851
r = 0.632
r = 0.731
miR-205
miR-296
let-7e
miR-193b
miR-148b
miR-196b
miR-155
C
miR-205 expression
miR-205 expression, RQ
PC3
PC3
PC3_BL
PC3_BL
BL_Xp114q21
BL_Xp114q21
D
miR-205 luciferase reporter assay
Luciferase Activity
miR205
miR205 +
miR20b +
miR375
RWPE 1 - Negative Luciferase control
RWPE 1/GFP - Control Transfected
RWPE 1/205 - Single Transfection
RWPE 1/20B, 375, 205 - Triple Transfection


E
miR-20b expression
miR-20b expression, RQ
0
5
10
15
20
25
30
35
RWPE1
LNCaP
LN3
Pro5
C4-2
C4-2B
F
miR-375 expresion
miR-20b expression, RQ
0
1000
2000
3000
4000
5000
6000
RWPE1
LNCaP
LN3
Pro5
C4-2
C4-2B
G
miR-375 expression
miR-expression, RQ
0
50
100
150
200
250
300
350
400
450
RWPE1
LNCaP
C4-2
C4-2B
H
Increased hsa-mir-205 expression in human prostate adenocarcinomas
Prostate adenocarcinomas
p = 0.029
miR-205 expression
Increased hsa-miR-205 expression in hESC
hESCs
p = 4.03116E-09
Committed and differentiating progenies

I


NALP1-locus SNP-RNA-regulated 'stemness' microRNAs enhance colony formation in agar by RWPE1 normal prostate epithelial cells
Colony-forming efficiency, %
0
1
2
3
4
5
6
7
8
9
10
* P < 0.05
*
*
*
*
RWPE1 EGFP
RWPE1 miR-20b
RWPE1 miR-375
RWPE1 miR-205
RWPE1 miR-20b/375
RWPE1 miR-20b/375/205